



Encodage neuronal des sons de parole : développements méthodologiques, générateurs neuronaux et application au malentendant appareillé

Ludovic Bellier

► To cite this version:

Ludovic Bellier. Encodage neuronal des sons de parole : développements méthodologiques, générateurs neuronaux et application au malentendant appareillé. Neurosciences [q-bio.NC]. Université Claude Bernard - Lyon I, 2015. Français. NNT : 2015LYO10135 . tel-01239484

HAL Id: tel-01239484

<https://theses.hal.science/tel-01239484>

Submitted on 7 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
Mention **NEUROSCIENCES**

Délivrée par
L'UNIVERSITÉ CLAUDE BERNARD LYON 1
ÉCOLE DOCTORALE *NEUROSCIENCES ET COGNITION* (ED476)

Présentée par **Ludovic Bellier**

En vue de l'obtention du grade de **Docteur en Neurosciences de l'Université de Lyon**

Intitulée

*Encodage neuronal des sons de parole :
développements méthodologiques, générateurs
neuronaux et application au malentendant appareillé*

Et soutenue publiquement le **25 septembre 2015**

Devant le Jury composé de

Dr. Marie Gomot en qualité de **Rapportrice**

Pr. Paul Avan en qualité de **Rapporteur**

Pr. Anne Didier en qualité d'**Examinatrice**

Dr. Daniele Schön en qualité d'**Examineur**

Dr. Anne Caclin en qualité de **Co-encadrante de thèse**

Pr. Hung Thai-Van en qualité de **Directeur de thèse**

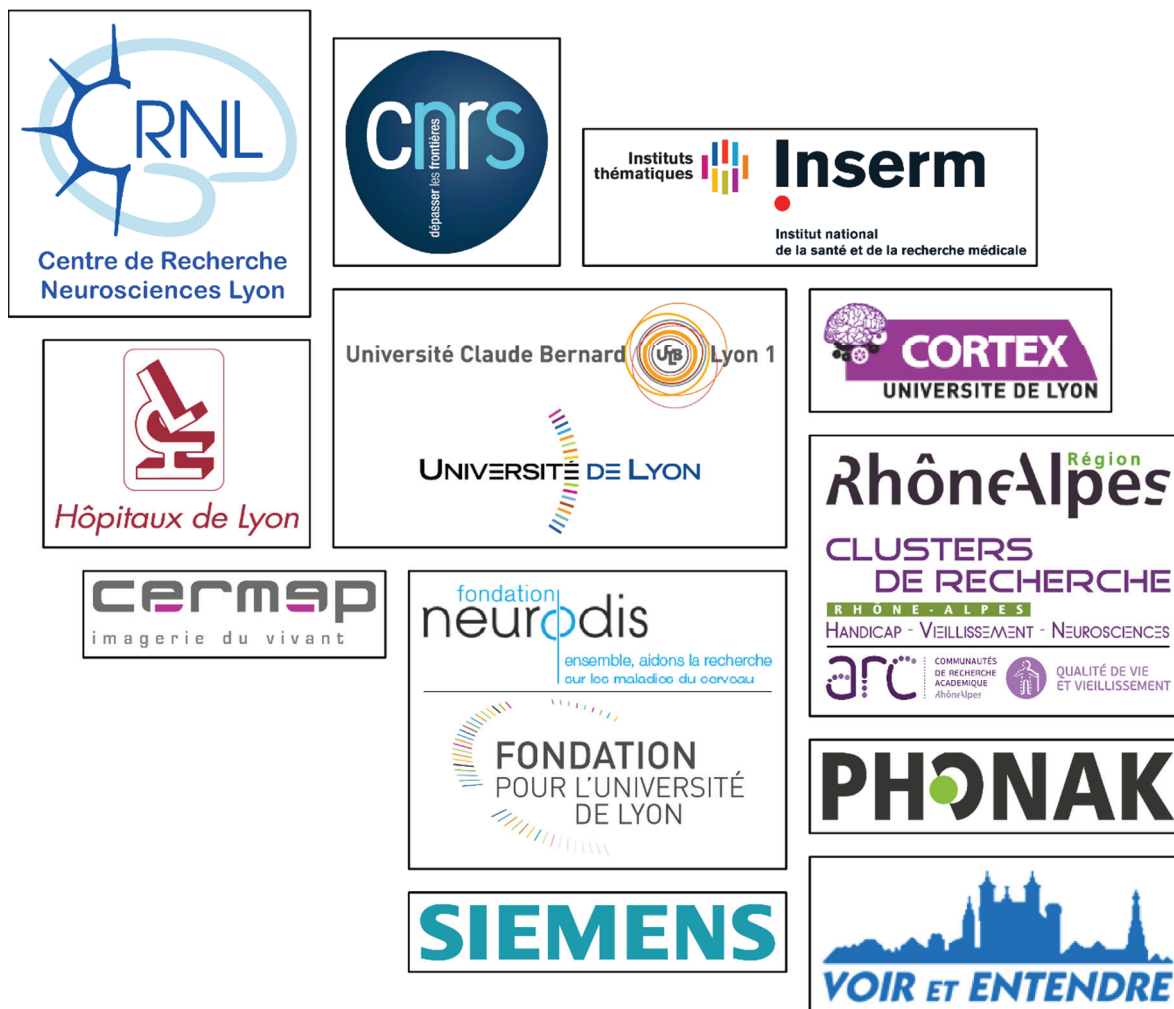
A Tania

«Tant qu'on n'aura pas diffusé très largement à travers les hommes de cette planète la façon dont fonctionne leur cerveau, la façon dont ils l'utilisent et tant que l'on n'aura pas dit que jusqu'ici que cela a toujours été pour dominer l'autre, il y a peu de chance qu'il y ait quoi que ce soit qui change.»

Henri Laborit, *Mon oncle d'Amérique*, 1980

Réalisé par Alain Resnais

Remerciements officiels



Ce travail de longue haleine, puisque commencé en septembre 2010, n'aurait pu se dérouler dans d'aussi bonnes conditions sans le concours bienveillant de précieux partenaires, auxquels je tiens ici à rendre hommage.

De par mes deux directeurs de thèse, Pr. Hung Thai-Van et Dr. Anne Caclin, j'ai d'abord eu la chance de naviguer alternativement entre les univers hospitalier et d'un laboratoire de recherche, respectivement dans le Service *Audiologie et Explorations oro-faciales* de l'Hôpital Edouard Herriot (Hospices Civils de Lyon), et dans l'équipe *Dynamique Cérébrale et Cognition* (DYCOG ; composante du LabEx CORTEX) du Centre de Recherche en Neurosciences de Lyon (CRNL ; tutelles CNRS, Inserm, Université Claude Bernard Lyon 1). Je garderai un souvenir quasi nostalgique de la découverte de ces deux environnements bien distincts, tout comme de l'accueil chaleureux et de l'aide inconditionnelle qui m'y ont été réservés.

En 2012 sont venus s'ajouter deux nouveaux terrains de recherche (littéralement) à mon espace de travail : le Service *Neurologie fonctionnelle et Épileptologie* de l'Hôpital Neurologique Pierre Wertheimer, pour les enregistrements de Stéréo-Encéphalographie (SEEG) chez le patient épileptique implanté (axe « Générateurs neuronaux »), et le laboratoire de l'audition *Voir et Entendre* (axe « Malentendant appareillé »). J'ai notamment eu l'immense privilège d'assister, pendant 2 mois, au quotidien des audioprothésistes de l'équipe de Jean-François Vesson. A nouveau, deux découvertes mémorables, avec un contact cette fois-ci plus direct avec les patients, et en conséquence un très grand enrichissement humain.

La liste des précieux partenaires ne serait pas complète sans mentionner les deux partenariats industriels nous ayant permis d'avancer très significativement le long de l'axe « Malentendant appareillé ». D'abord avec les allemands de chez Siemens en 2012 et 2013, pour la mise au point d'une méthodologie de stimulation au travers des prothèses auditives, puis avec les suisses de chez Phonak en 2014 et 2015, pour l'étude longitudinale chez le malentendant appareillé. Ces deux collaborations auront été l'occasion de découvrir avec admiration toute la réactivité et les moyens que mettent ces sociétés à améliorer le bien-être quotidien de millions de malentendants.

Concernant les financements, j'ai bénéficié durant mes trois premières années de thèse d'une bourse de la Région Rhône-Alpes sous gestion du Cluster 11 *Handicap-Vieillesse-Neurosciences*, auquel succéda en 2012 la Communauté de Recherche Académique ARC 2 *Qualité de vie et Vieillesse*. La quatrième année a ensuite été financée sur des crédits de mes deux directeurs de thèse, sous gestion de l'Inserm, puis du Cermep. Enfin, un partenariat industriel a été établi avec l'entreprise suisse de manufacture d'appareils auditifs Phonak, pour le financement de ma cinquième et dernière année de thèse. Les fonds versés ont été pilotés par la Fondation Neurodis, abritée par la Fondation Pour l'Université de Lyon depuis le 30 septembre 2014. J'aimerais alors remercier les décisionnaires de ces différentes entités pour leur confiance et leur soutien, ainsi que les gestionnaires pour leur efficacité et leur sympathie.

Remerciements (un peu moins solennels)

Ces cinq années de thèse auront été le lieu d'un enrichissement scientifique et humain bien au-delà de l'imaginable, qui aurait sans aucun doute été bien moindre sans l'intervention de multiples personnes, d'horizons divers. Je tiens ici à leur adresser mes plus sincères remerciements (et mes plus plates excuses à ceux que j'oublierais).

A Anne Caclin et Hung Thai-Van, mes deux directeurs de thèse, pour toute leur confiance et leur soutien inconditionnel, pour tous leurs précieux et judicieux conseils, ainsi que pour m'avoir offert un accès privilégié aux mondes de la recherche, de la clinique et de l'industrie, pour nos réunions « 9h30 - croissants » dont je garderai longtemps le goût, et pour leur profonde et sincère amitié développée au fil des années.

A Olivier Bertrand, directeur du CRNL et de cette incroyable équipe qu'est DYCOG, pour son accueil bienveillant et son intuition, qui aura potentialisé notre détour par le service épilepto' de l'Hôpital Neurologique de Lyon Est avec les résultats que l'on connaît. A ce propos, j'en profite pour aussi adresser mes remerciements à Aurélie Bidet-Caulet, Philippe Albouy et Mathilde Petton pour leur très appréciée collaboration dans ce projet.

A Evelyne Veuillet, Michel Mazzuca, Rafael Laboissière, Fabien Millioz et tous les membres passés et présents de l'équipe PACS, au pavillon U de l'Hôpital Edouard Herriot. Dans l'équipe DYCOG, à Doc' Bouet, Manu, Françoise et Jérémie pour les rires échangés durant nos sessions de travail ; à mes frères Philippe et Gaëtan (paraît-il) et pourquoi pas mes sœurs Margaux et Suzanne ; à Aline, Perrine, Yohana et Laurie-Anne pour nos tours de piste du Vinatier ; à Karim, Jean-Philippe et Antoine pour les ateliers foot-frisbee, ping-pong et méditation ; mention spéciale à Anatole pour m'avoir fait découvrir le *Lydian Chromatic Concept of Tonal Organization* et les lunettes à trous ; à Patrick, PierreM et Thibaut pour leur assistance sans faille ; et enfin à tous les étudiants et chercheurs qui ont été pour moi comme une seconde famille.

Aux personnels de HEH et du Neuro, pour leur accueil parfaitement hospitalier. A Jean-François Vesson et sa formidable équipe d'audioprothésistes, pour leur passion communicative.

A tous les patients malentendants, épileptiques, et tous les autres participants aux différentes études de cette thèse, pour s'être prêtés au jeu d'écouter une même syllabe /ba/ pendant toute une heure (moins pour patients épileptiques, rassurons-nous), et pour de superbes moments emplis d'humanité.

A mes amis de Lyon et de Paris, aux côtés desquels j'ai eu la chance de grandir : Philippe, Mauricio, Romain, Noé, et à toute la clique de Voisins-le-Bretonneux. Enfin à mes parents Geneviève et Jacques et à ma sœur Cécile, pour leur soutien et leur amour depuis toujours. Et en parlant d'amour, surtout, surtout à Tania, pour tout son Être !

Résumé

La Speech ABR (pour Speech Auditory Brainstem Response, ou Réponse du tronc cérébral auditif aux sons de parole) est un marqueur électrophysiologique de l'encodage des sons de parole dans le système nerveux auditif. Depuis sa découverte au début des années 90, la Speech ABR a été utilisée dans un nombre croissant d'études touchant à plusieurs domaines des Neurosciences (musique, langage, vieillissement, autisme, troubles des apprentissages, attention sélective, entre autres). Cependant, la connaissance des générateurs neuronaux de cette réponse électrophysiologique, qui ne repose à ce jour que sur un faisceau de preuves relativement faibles et potentiellement biaisées, n'a que peu évolué, tout comme la méthodologie de recueil et de stimulation. Pourtant, des verrous techniques tels que l'usage exclusif d'un type d'écouteurs particuliers pour délivrer les sons de parole qui, s'ils permettent de limiter la contamination de la réponse évoquée par un artéfact de stimulation, n'en présentent pas moins d'importantes limitations en termes de niveau de stimulation, empêchent l'utilisation de la Speech ABR chez des sujets avec des troubles de l'audition. Or, de par sa représentation fidèle des propriétés spectro-temporelles de la parole, la Speech ABR apparaît aujourd'hui comme un outil très prometteur pour caractériser la nature et suivre la remédiation des troubles de l'audition, problème de santé publique sans cesse croissant.

Les études constituant ce travail de thèse se sont articulées autour de deux axes : la question des générateurs neuronaux de la Speech ABR, et l'usage de la Speech ABR chez le malentendant appareillé. Pour chacun de ces deux axes, des innovations méthodologiques ont précédé et permis l'investigation des questions posées.

Concernant la question des générateurs neuronaux, nous avons d'abord développé une méthodologie de recueil simultané, en électroencéphalographie (EEG) 32 voies, de la Speech ABR et des Potentiels Evoqués Auditifs corticaux (PEA corticaux ; étude 1). Les topographies de la Speech ABR ainsi obtenues, et jusque-là inédites, ont suggéré l'existence d'une participation corticale à la génération de cette réponse, alors qu'elle avait toujours été décrite comme strictement sous-corticale. Afin d'investiguer plus précisément cette hypothèse, nous avons étudié l'activité cérébrale évoquée par les mêmes sons de parole directement à l'intérieur du cortex auditif de patients épileptiques implantés avec des électrodes de Stéréo-EEG (SEEG). Ces enregistrements intracrâniens ont confirmé l'existence d'une activité Speech ABR au niveau cortical, apparemment restreinte au cortex auditif primaire de façon bilatérale (étude 2). Ce résultat apporte un éclairage nouveau à la fois sur l'interprétation de la nature de la Speech ABR, et sur notre connaissance des capacités de codage temporel du cortex auditif humain.

Pour préparer l'utilisation de la Speech ABR chez le patient malentendant, nous avons développé une méthodologie innovante de stimulation sonore directement *via* les appareils de correction auditive, permettant d'atteindre une très haute intensité de stimulation tout en conservant une qualité de restitution sonore inégalable (étude 3). Cette avancée a permis la

mise en œuvre d’une étude longitudinale posant la question de la plasticité neuronale induite par le port d’aides auditives chez le malentendant presbyacousique, associant mesures comportementales et électrophysiologiques. Les résultats préliminaires vont dans le sens d’une évolution, au cours des 4 premiers mois d’appareillage, des différentes mesures vers les valeurs normo-entendantes, et objectivent de ce fait l’influence des appareils auditifs sur la plasticité neuronale de différentes étapes du système nerveux auditif (étude 4).

Mots-clés

Neurosciences ; électrophysiologie ; audition ; parole ; encodage neuronal ; tronc cérébral auditif ; cortex auditif ; *Speech Auditory Brainstem Response* (*Speech ABR*) ; *Frequency-Following Response* (*FFR*) ; *Potentiels Evoqués Auditifs corticaux* (*PEA corticaux*) ; électroencéphalographie de scalp (EEG) ; stéréo-électroencéphalographie (SEEG) ; méthodologie ; traitement de données ; générateurs neuronaux ; malentendant ; presbyacousie ; appareils auditifs ; plasticité neuronale.

Laboratoire de rattachement

Centre de Recherche en Neurosciences de Lyon (CRNL), équipe *Dynamique cérébrale et Cognition* (DYCOG), Inserm U1028 - CNRS UMR 5292 - Université Claude Bernard Lyon 1 ; en deux lieux :

- Centre hospitalier Le Vinatier, bâtiment 452, 95 boulevard Pinel 69500 Bron
- Hôpital Edouard Herriot (Hospices Civils de Lyon), service *Audiologie et Explorations oro-faciales*, 5 place d’Arsonval 69003 Lyon

Abstract

Speech ABR (for Speech Auditory Brainstem Response) is an electrophysiological marker of speech sound encoding in the auditory nervous system. Since its discovery in early 90s, Speech ABR has been used in a growing number of studies related to several fields in Neuroscience (music, language, aging, autism, learning disorders, selective attention, among others). However, knowledge of the neural generators of this electrophysiological response, to date on only based a relatively weak and potentially biased body of evidence, has only changed little, just as the recording and stimulation methodology. Yet, technical obstacles such as the exclusive use of a specific kind of earphones to deliver speech sounds which, if avoiding contamination of the evoked response by a stimulus artifact, inherently exhibit important limitations in terms of stimulation sound level, prevents the use of Speech ABR in participants with hearing impairments. Now, with its faithful representation of speech spectro-temporal characteristics, the Speech ABR today appears as a very promising tool to characterize the nature and monitor the remediation of hearing disorders, an ever increasing public health problem.

Studies constituting this thesis work are articulated around two axes: the question of Speech ABR's underlying neural generators, and the use of Speech ABR in hearing aid users. For each of these two axes, methodological innovations preceded and enabled investigation of the raised questions.

Regarding Speech ABR's underlying neural generators, we first developed a methodology for simultaneous recording, using 32-channel electroencephalography (EEG), of both Speech ABR and cortical Auditory Evoked Potentials (cortical AEP; study 1). Obtained topographies of Speech ABR, until then unpublished, suggested the existence of a cortical participation to the generation of this response, whereas it always had been described as strictly subcortical. To investigate more precisely this hypothesis, we studied cerebral activity evoked by the same speech sounds directly within the auditory cortex of epileptic patients implanted with Stereo-EEG electrodes (SEEG). Intracranial recordings revealed the existence of a cortical *Speech ABR* activity, apparently restricted to bilateral primary auditory cortex (study 2). This result sheds new light on both the interpretation of Speech ABR's nature, and our knowledge of temporal coding capabilities of the human auditory cortex.

To prepare for the use of Speech ABR in hearing-impaired patients, we developed an innovative methodology of sound stimulation directly *via* hearing aids, allowing to reach very high stimulation intensities while maintaining an unmatched quality of sound restitution (study 3). This advance enabled the design of a longitudinal study assessing neural plasticity induced by the use of hearing aids in hearing-impaired patients with presbycusis, combining behavioral and electrophysiological measures. Preliminary results suggest an evolution, over the first four months of hearing aid use, of the different measures toward normal-hearing

values, thus objectifying the influence of hearing aids on neuronal plasticity within different stages of the auditory nervous system (study 4).

Keywords

Neuroscience; electrophysiology; hearing; speech; neural encoding; auditory brainstem; auditory cortex; *Speech Auditory Brainstem Response* (*Speech ABR*); *Frequency-Following Response* (*FFR*); *cortical Auditory Evoked Potentials* (*cortical AEP*); scalp-recorded electroencephalography (EEG); stereo-electroencephalography (SEEG); methodology; data processing; neural generators; hearing impaired; presbycusis; hearing aid; neural plasticity.

Table des matières

1. Préambule	1
 Première partie - Contexte théorique	 4
 2. La parole	 5
2.1. Contextualisation, ou <i>petite histoire de la parole</i>	5
2.1.1. Concept d'information	5
2.1.2. La communication à travers les espèces	6
2.1.3. La parole dans l'humanité	8
2.2. La parole : définition et structure	10
2.2.1. Etymologie et dérive phonétique	10
2.2.2. La parole en poupée russe	10
2.2.3. Composition des phonèmes	11
2.2.4. Information paralinguistique	13
2.3. Le cycle de la parole	13
2.3.1. Production	15
<u>2.3.1.1. La théorie source-filtre</u>	15
<u>2.3.1.2. Boucle audio-phonatoire</u>	18
<u>2.3.1.3. Phonétique articulatoire</u>	18
<u>2.3.1.4. Parole naturelle ou synthétique</u>	21
<u>2.3.1.5. Techniques de synthèse</u>	22
2.3.2. Transmission	23
<u>2.3.2.1. Notion de bruit</u>	23
<u>2.3.2.2. Stratégies d'adaptation au bruit</u>	25
2.3.3. Perception	26
<u>2.3.3.1. Modalités sensorielles</u>	26
<u>2.3.3.2. Le système auditif humain</u>	28
<i>L'oreille</i>	28
<i>Oreille externe</i>	29
<i>Oreille moyenne</i>	30
<i>Oreille interne - anatomie</i>	32
<i>Oreille interne - physiologie</i>	36

<i>Voies nerveuses auditives sous-corticales</i>	38
<i>Thalamus et cortex auditif</i>	40
2.3.4. Cognition	41
2.3.5. Boucler la boucle	43
2.4. La parole et ses imperfections	43
2.5. Pathologies liées et traitements associés	45
2.5.1. Surdités de transmission	46
2.5.2. Surdités de perception - Zoom sur la presbyacousie	46
3. Electrophysiologie de l'audition et autres outils de mesure	49
3.1. Techniques à disposition du clinicien et du chercheur	49
3.1.1. Vue d'ensemble	49
3.1.2. Outils d'exploration clinique des voies auditives	50
<u>3.1.2.1. Comportement</u>	50
<u>3.1.2.2. Anatomo-physiologie</u>	53
3.1.3. Electrophysiologie de l'audition.....	53
3.2. Speech ABR	57
3.2.1. Méthodologie	58
<u>3.2.1.1. Stimulation</u>	59
<i>Stimulus</i>	59
<i>Stimulateur</i>	61
<i>Paramètres de stimulation</i>	62
<u>3.2.1.2. Recueil</u>	66
<u>3.2.1.3. Traitements et analyses</u>	69
<i>Traitements</i>	69
<i>Analyses</i>	72
3.2.2. Générateurs neuronaux	75
<u>3.2.2.1. Etat des connaissances</u>	75
<u>3.2.2.1. Click ABR</u>	77
<u>3.2.2.2. FFR</u>	78
<u>3.2.2.3. Cortex auditif ?</u>	78
3.2.3. Champs d'application	80
<u>3.2.3.1. Les débuts d'un marqueur prometteur</u>	80
<u>3.2.3.2. Audition</u>	81
<i>Asymétrie du système auditif</i>	81
<i>Parole et phonétique</i>	82
<i>Parole dans le bruit</i>	82
<i>Entraînement auditif et plasticité neuronale</i>	83
<i>Développement</i>	84

<i>Vieillessement</i>	84
3.2.3.3. <u>Expertises auditives</u>	86
<i>Langage</i>	86
<i>Musique</i>	88
3.2.3.4. <u>Déficiences auditives et troubles liés</u>	90
<i>Troubles de l'audition</i>	90
<i>Appareillage auditif</i>	92
<i>Trouble des apprentissages</i>	92
3.2.3.5. <u>Cognition</u>	94
<i>Perception catégorielle</i>	94
<i>Effet de transformation verbale</i>	94
<i>Attention sélective</i>	94
<i>Emotions</i>	95
<i>Interactions audiovisuelles</i>	95
3.2.3.6. <u>Pathologies</u>	95
<i>Amusie congénitale</i>	95
<i>Cécité</i>	96
<i>Epilepsie</i>	96
<i>Hyperactivité</i>	96
<i>Schizophrénie</i>	96
<i>Troubles du spectre autistique</i>	96
3.2.3.7. <u>Connaissances générales sur la Speech ABR</u>	98
<i>Effet du genre</i>	98
<i>Facteurs génétiques et environnementaux</i>	98
<i>Lien à d'autres mesures (MMN, OAE)</i>	98
3.2.4. <u>Etat des lieux</u>	99

Deuxième partie - Travaux expérimentaux 103

4. Axe « Générateurs neuronaux de la Speech ABR »	105
4.1. Etude 1 : Topographie de la Speech ABR	106
4.2. Etude 2 : Composante corticale de la FFR	114
5. Axe « Speech ABR chez le malentendant appareillé »	133
5.1. Etude 3 : Appareils auditifs dans un protocole Speech ABR	134
5.2. Etude 4 : Plasticité neuronale chez le malentendant appareillé	142

Troisième partie - Discussion générale	189
6. Discussion générale	191
6.1. Résumé des résultats	191
6.2. Discussion, limites et perspectives	192
6.2.1. Méthodologie de la Speech ABR	192
6.2.2. Générateurs neuronaux de la Speech ABR	193
6.2.3. Plasticité neuronale chez le malentendant appareillé	195
Références bibliographiques	197

1. Préambule

L'Homme est un animal social, disait Aristote. Une traduction qui rendrait peut-être davantage honneur à sa pensée changerait *social* pour *politique*. Dans les deux cas, ce postulat illustre bien le quotidien de la plupart des humains : nous vivons en groupes sociaux, et passons un temps plus ou moins important, selon la sensibilité de chacun, à échanger sur nos pensées, émotions et ressentis. Ces nombreux échanges utilisent préférentiellement l'outil merveilleux qu'est la parole, vectrice de communication entre nos esprits.

Mais si notre utilisation quotidienne de la parole tendrait à nous faire percevoir ce phénomène comme aisé et automatique, c'est parce que nous sommes normalement, sans forcément en avoir conscience, des experts dans le maniement de cet extraordinaire instrument qu'est l'appareil phonatoire, et dans la reconnaissance des signaux produits. La participation effective au fourmillement langagier humain nécessite en réalité la possession et le bon fonctionnement de nombre de facultés et organes sous-jacents : appareil phonatoire et contrôle moteur pour la prononciation, système auditif pour la perception, compétences attentionnelles et cognitives pour l'interprétation, la réflexion et la décision éventuelle de produire une réponse. En référence à la *boucle perception-action* bien connue des neuroscientifiques, j'ai choisi de donner le nom de *cycle de la parole* à cet ensemble d'étapes.

Les travaux composant cette thèse sont centrés sur l'encodage neuronal des sons de parole dans le système auditif humain, plus particulièrement au niveau sous-cortical. Cette dernière précision était une de nos hypothèses de travail, que nous allons quelque peu malmenier au fur et à mesure des pages du présent manuscrit. Elle vient de la nature classiquement acceptée de l'objet d'étude au centre de cette thèse : la *Speech Auditory Brainstem Response* (Speech ABR), ou Réponse du tronc cérébral auditif aux sons de parole.

Du fait de sa grande fidélité au *stimulus* utilisé pour l'évoquer, c'est-à-dire plus simplement de sa ressemblance au son de parole délivré, cette réponse électrophysiologique constituerait un candidat idéal pour étudier les troubles de l'audition de patients malentendants si seulement les outils disponibles pour restituer au patient les *stimuli* d'intérêt permettaient d'atteindre une intensité de stimulation suffisante tout en conservant une qualité de restitution sonore acceptable.

Ces deux thématiques, nommément les « générateurs neuronaux de la Speech ABR », et « l'utilisation de la Speech ABR chez le patient malentendant », ont été les deux axes directeurs des études réalisées pendant ma thèse. Pour chacun de ces deux axes, un travail méthodologique a précédé l'investigation scientifique à proprement parler.

Ainsi, pour l'axe « Générateurs neuronaux », nous avons d'abord développé une méthodologie de recueil topographique simultané de la Speech ABR et des *Potentiels Evoqués Auditifs corticaux* (PEA corticaux). L'étude des topographies alors obtenues nous a incités à poursuivre l'investigation en changeant de technique de mesure pour passer du scalp au cortex auditif *in situ*, grâce à la stéréo-EEG (SEEG) où des patients épileptiques sont implantés avec des électrodes intracrâniennes dans le but de localiser le foyer d'origine de leurs crises, avant résection potentielle de la zone impliquée.

Concernant l'axe « Malentendant appareillé », nous avons mis au point une méthodologie de stimulation *via* les appareils auditifs, qui combine très forte intensité et excellente restitution sonore, innovante dans le cadre du protocole d'enregistrement de la Speech ABR. Puis nous avons appliqué cette méthode pour poser la question de la plasticité neuronale induite par le port d'appareils auditifs chez le patient presbyacousique.

Pour comprendre pleinement le récit des différentes études présentées ici, la possession de notions préalables est vivement souhaitable, lesquelles seront justement présentées dans la partie « Contexte théorique » de ce présent manuscrit.

Dans un premier chapitre (2.), nous décrirons l'objet de mesure de l'objet d'étude de cette thèse (respectivement la parole et la Speech ABR) dans son contexte, son utilité individuelle et sociale, et sa nature. Comprendre l'encodage de la parole dans le système nerveux humain nécessite la connaissance de ce qu'est la parole (sous-chapitres 2.1. et 2.2.), des différentes informations qu'elle contient et véhicule. Puis nous suivrons le *cycle de la parole* (2.3.) pour aborder les notions de production, perception, et cognition. Ce premier chapitre se terminera en explorant les imperfections de ce cycle (2.4.) et les pathologies associées (2.5.).

Dans un second chapitre (3.), nous nous intéresserons aux différents outils de mesure du système auditif humain (3.1.), puis nous mettrons la Speech ABR en exergue (3.2.). Nous parlerons alors de ses méthodologies d'acquisition et de traitement de signal (3.2.1.), de ses générateurs (3.2.2.), et de ses champs d'applications divers (3.2.3.).

Ce plan en deux temps explorera donc en premier le phénomène qu'est la parole (avec les organes et fonctions impliqués), et en second la mesure scientifique d'une sous-partie de ce phénomène, à savoir la perception des sons de parole.

Première partie

Contexte théorique

2. La parole

2.1. Contextualisation, ou *petite histoire de la parole*

2.2.1. Concept d'information

Si la nature de l'Univers dans lequel nous vivons défie encore largement l'entendement humain, des théoriciens de la physique quantique ont récemment avancé l'*information* comme constituant fondamental, plutôt que matière et énergie comme jusqu'alors admis (Vedral, 2010 ; Davies & Gregersen, 2014). Ce qui est sûr, c'est que l'information circule dans et entre tous les niveaux d'organisation du Vivant. Et dans le monde socio-technologique que nous nous sommes construit, ne dit-on pas que nous vivons au Siècle de l'information ?

Dérivé du latin *informo*, l'information est au sens le plus brut « l'action de donner forme à quelque chose ». Nous sommes là bien loin des significations plus usuelles de ce mot, rapprochant information de *connaissance* ou d'*actualité* (sauf à considérer que le but des fameuses chaînes d'information serait de *donner forme* à nos esprits). Dans sa *Nouvelle grille*, le neurobiologiste français Henri Laborit reprend la définition du mathématicien américain Norbert Wiener, père notoire de la cybernétique : « L'information n'est ni masse, ni énergie ; l'information est information » (Wiener, 1961), à laquelle il ajoute la distinction entre *information-structure* et *information circulante* (Figure 2.1. ; Laborit, 1985). La citation de Wiener peut se comprendre à l'aide d'un exemple très simple : en considérant un mot imprimé, comme « relapo », et en changeant l'ordre de ses lettres, pour « parole », ni la masse ni l'énergie ne diffèrent entre les deux mots : seule l'information naît de l'organisation particulière des lettres du second.

L'information-structure est à rapprocher du sens étymologique premier du mot information, en cela qu'elle représente le système de relations qui, appliquées aux éléments unitaires d'un ensemble, lui donnent sa forme, sa *structure*. Notons que cela s'observe à toutes les échelles de l'Univers, à tous les *niveaux d'organisation* tels qu'introduits par le même Laborit (sauf peut-être au niveau des *particules élémentaires*, s'il ne s'agit de l'information elle-même). Concrètement, l'information-structure est ce qui lie les atomes entre eux pour former les molécules, ce qui associe les organites cellulaires pour former les cellules, ce qui agence les organes pour former les animaux, ce qui organise les humains pour former la société, ce qui agence les planètes pour former les galaxies, *et cetera*. Selon les mots de Laborit, c'est cette information-structure qui nous « permet de distinguer un homme d'un éléphant ».

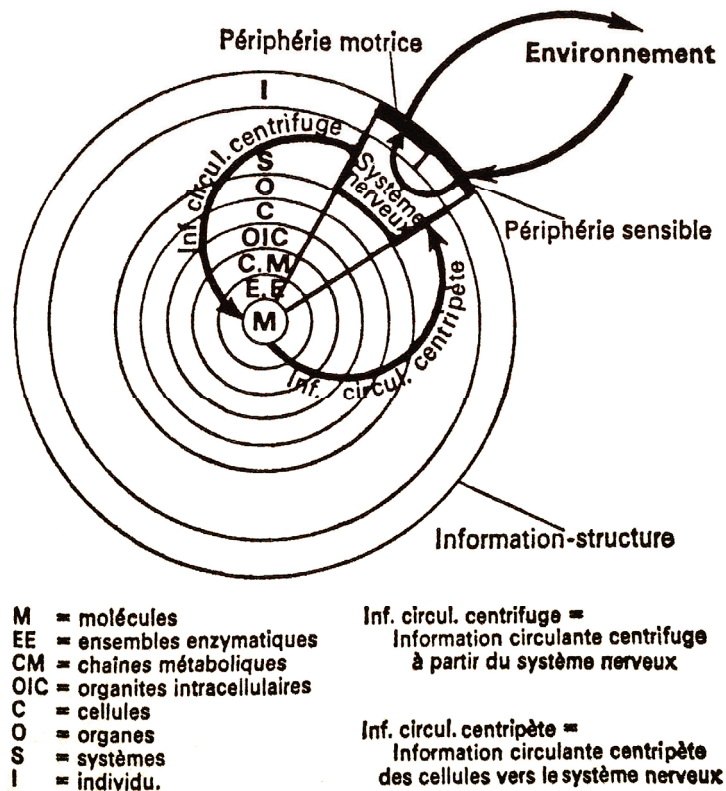


Figure 2.1. Information-structure et information-circulante pour des niveaux d'organisation des molécules à l'individu (adapté de *La nouvelle grille*, Laborit, 1985).

L'information circulante se rapproche quant à elle du sens courant du mot information, en désignant les messages transmis à l'intérieur d'un même niveau d'organisation, ou entre différents niveaux. Un électron passant d'un atome à un autre, des neurotransmetteurs se fixant sur des récepteurs post-synaptiques, une phrase tombant dans l'oreille d'un interlocuteur, sont autant d'exemples de ce concept. Notons au passage le lien entre information-structure et information circulante : chaque vecteur d'information circulante est lui-même modelé par une information-structure, et peut modifier la forme d'un élément récepteur pour incarner de ce fait l'information-structure.

2.1.2. La communication à travers les espèces

Dans le monde du Vivant, chaque organisme, qu'il soit uni- ou pluricellulaire, possède une information-structure propre à chacun de ses niveaux d'organisation (molécules, organites cellulaires, cellules, organes, organisme), leur permettant de maintenir leur forme et donc la structure de l'organisme ainsi composé. L'information circulante permet d'assurer, à et entre tous ces niveaux, un équilibre, une *homéostasie*, un « dialogue biologique ».

En utilisant cette grille de lecture, la communication peut se définir comme l'échange d'information circulante entre différents organismes d'une même espèce, voire d'espèces différentes, ce qui reste conforme à son étymologie latine signifiant « action de mettre en commun ». Et les exemples à travers le Vivant sont nombreux et variés : des phéromones de la fourmi au chant des baleines, en passant par les couleurs du caméléon et la grimace d'un chien menaçant, force est de constater que toutes les modalités sensorielles sont sollicitées. Dans tous les cas, ces échanges d'information circulante permettent une harmonisation des organismes entre eux, et contribuent ainsi à leur pérennité, soit au maintien de leur information-structure.

L'information circulante passe aussi parfois la barrière des espèces. Un exemple attendrissant provient d'une étude récente ayant montré qu'un long regard entre un chien et son maître engendrait, chez le maître comme chez son chien, une augmentation du taux d'ocytocine, hormone liée à l'empathie et à l'attachement (Nagasawa *et al.*, 2015). Mais ce cas de figure nécessite de partager un même mode de communication (ici le regard), ce qui empêche nombre d'interactions inter-espèces. Dans une initiative pour contourner cette limite, Diana Reiss (chercheuse en sciences cognitives et spécialiste des dauphins), Peter Gabriel (chanteur de Genesis), Neil Gershenfeld (créateur du concept *Fab lab* et professeur au MIT) et Vint Cerf (un des pères d'Internet) ont exposé dans une conférence TED de 2013 leur projet de créer un Internet des animaux, en inventant pour cela des interfaces que les animaux pourraient contrôler pour interagir avec d'autres espèces, dont les humains (disponible à l'adresse <http://interspecies-internet.org>). Le flux d'information circulante ainsi établi pourrait alors permettre à tout un chacun de prendre conscience des capacités émotionnelles et intellectuelles des animaux (au moins des grands mammifères puisque l'initiative concerne les dauphins, les singes et les éléphants).

Si la communication existe dans toutes les modalités sensorielles, l'audition permet de communiquer des signaux extrêmement fins et modulables, et donc d'exprimer une palette d'émotions ou d'états mentaux quasi infinie. Beaucoup d'animaux l'utilisent, souvent relativement grossièrement, tout du moins en apparence. Certaines classes ou espèces se distinguent par la finesse étonnante des sons produits et un répertoire digne d'une langue. C'est par exemple le cas des oiseaux chanteurs, capables de produire des mélodies de façon innée, et d'apprendre les chants de ses congénères avec des variations selon les zones géographiques comparables aux dialectes humains (Xing *et al.*, 2013). Aussi, et comme j'ai pu le découvrir lors de mon stage de M1 dans l'équipe de Sylvie Granon (Université Paris-Sud 11, Orsay), les souris sont capables de produire des vocalisations ultrasoniques, variables selon le contexte, et composées de syllabes structurées et redondantes (Chabout *et al.*, 2012 ; Figure 2.2.). Néanmoins, ce répertoire de productions vocaliques reste loin de la complexité du langage oral humain.

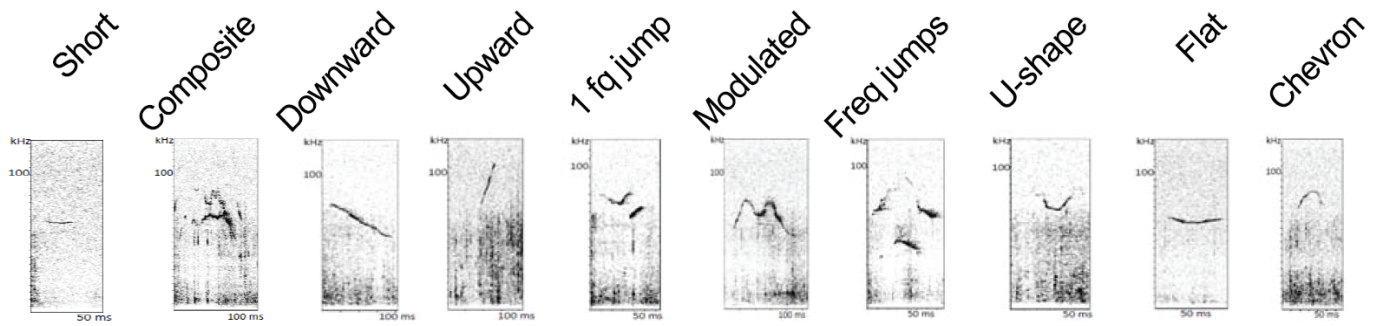


Figure 2.2. Répertoire des vocalisations ultrasoniques de la souris mâle adulte (adapté de *Chabout et al.*, 2012).

L'espèce humaine utilise ainsi une grande variété de sons langagiers, associée à une combinatoire vertigineuse qui place la parole comme l'un des modes de communication auditive les plus complexes du Vivant. Mais complexité n'est pas unicité, et comme nous venons de le voir au long de ce voyage succinct à travers les niveaux d'organisation de l'Univers et du Vivant, la parole n'est qu'une des formes d'information circulante parmi une infinité de modes d'échanges existants. En restant dans l'espèce humaine, de nombreuses de façons de communiquer, consciemment ou non, existent autres que la parole, et sont regroupés sous l'étiquette de *communication non-verbale* : expressions faciales, regard, gestes, attitudes, codes vestimentaires, contacts, rires, pleurs, odeurs, arts, et même langues sifflées.

Enfin, si la parole est la manifestation de la pensée, son absence dans une espèce ne signifie pas fatalement que les membres de cette espèce soient dépourvus de pensées, mais simplement qu'ils ne possèdent pas l'appareil phonatoire requis ou le contrôle moteur nécessaire pour manier cet appareil. Nombre d'exemples issus de l'éthologie ont d'ailleurs montré que de nombreuses espèces possédaient des émotions, une conscience de soi, voire même une mémoire épisodique. C'est le cas des gorilles, avec la très médiatisée *Koko* qui, ayant appris plus de mille signes d'une langue des signes adaptée à sa morphologie auprès de la chercheuse américaine Penny Patterson (*Gorilla Sign Language* ; Patterson & Linden, 1981), est capable de raconter des souvenirs et, plus étonnant encore, de pleurer la mort de son ami Robin Williams.

2.1.3. La parole dans l'humanité

La parole, en tant que faculté à exprimer oralement la pensée, est intimement liée à la condition humaine. Nous en faisons l'usage quotidiennement, si bien qu'elle nous semble évidente, naturelle, et qu'il serait difficile d'imaginer ce que serait l'Humanité sans elle. Sa place centrale dans nos vies explique sans doute pourquoi elle a passionné les philosophes et scientifiques à travers les Âges, échelles d'observation et disciplines liées.

L'origine de la parole serait postérieure à l'acquisition d'une posture verticale bipède et à l'abaissement du larynx, offrant selon certaines hypothèses une plus grande liberté articulatoire, ces deux évolutions datant de plus d'un million d'années. Le long perfectionnement qui s'en suivit a progressivement rendu possible la transmission des savoirs et savoir-faire, soit de la culture. Suivant cette logique, le philosophe français Régis Debray a proposé le concept de *logosphère* pour décrire la période allant de la naissance de l'écriture (4000 ans avant J.-C.) à l'invention de l'imprimerie (aux alentours de 1450 après J.-C.), où les textes manuscrits et la lecture étaient réservés à une minorité de savants, et où l'oralité était donc le vecteur principal pour l'échange interindividuel de techniques et connaissances à travers les générations et les siècles (Debray, 1991). Notons que si les deux sphères suivantes, nommées *graphosphère* et *vidéosphère*, ont mis en exergue la transmission culturelle respectivement par le livre et la communication audiovisuelle, la parole possède toujours un rôle central dans nos vies du fait qu'elle soit directe, rapide, et très riche en information. Elle permet de penser, de simuler le monde, de créer des concepts. Sa capacité à transmettre des connaissances d'un individu à d'autres, comme par exemple la présence d'un ours à droite après la troisième cascade, ou les effets relativement néfastes des champignons à tête rouge à points blancs, a très probablement constitué un avantage évolutif majeur.

A l'échelle de l'individu, la parole permet de former le cerveau au langage par une simple exposition répétée (apprentissage implicite). Au cours de la vie, nous apprenons d'abord à dénommer certains objets de notre entourage, puis faisons ensuite abstraction des détails non-pertinents pour interagir avec ceux-ci, en formant des concepts. Cela se fait naturellement, dans un environnement donné et pour une langue donnée, soit dans une culture donnée.

Les linguistes et anthropologues américains Edward Sapir et son élève Benjamin Lee Whorf ont avancé dans les années 1950 une théorie de la relativité linguistique, aussi connue sous le nom d'hypothèse Sapir-Whorf, postulant que le langage influe sur notre façon de voir le monde (Hojjer, 1954). Dans la version la plus radicale de cette théorie, nous ne pourrions concevoir une réalité sans mots pour l'exprimer. Une étude de psychologie expérimentale menée en 2006 a supporté une version plus modérée de cette hypothèse, en montrant que les catégories sémantiques possédées pour désigner les couleurs influençaient leur discrimination, seulement lorsque les couleurs étaient présentées dans l'hémi-champ visuel droit (Gilbert *et al.*, 2006). Or, l'information provenant de cette partie du champ visuel est traitée dans l'hémisphère gauche du cerveau, où le langage est majoritairement traité chez les droitiers (Zatorre *et al.*, 1992 ; Hickok & Poeppel, 2007), ce qui suggère effectivement une influence des noms de couleurs sur la perception de celles-ci.

Aussi, la comparaison de langues différentes révèle souvent des mots sans équivalent traduisible, comme le sont pour le français le *leitmotiv* allemand (désignant une formule revenant régulièrement dans une œuvre littéraire ou musicale) ou l'adjectif anglais *cheap* (bon marché). Le recours à de tels concepts dans un raisonnement mental sera facilité par la possession de ces mots, qui jouent alors un rôle de raccourci sémantique, de facilitateur rendant la pensée plus fluide, directe et précise.

Ces quelques exemples permettent d'illustrer le rôle majeur que joue la parole, ici *information circulante*, chez l'Homme, aussi bien au niveau sociétal qu'à l'échelle de l'individu. Avant d'étudier les différentes capacités (motrices, perceptives, cognitives) qui sous-tendent le bon fonctionnement du *cycle de la parole*, nous allons maintenant nous intéresser à la structure même de la parole, à ce qu'elle est et à ce qui la constitue—en quelque sorte à son *information-structure*.

2.2. La parole : définition et structure

2.2.1. Etymologie et dérive phonétique

Dans ce présent manuscrit, le mot *parole* est uniquement utilisé pour désigner la manifestation orale articulée du langage verbal (humain, inévitablement). Son étymologie la plus lointaine remonte au grec ancien, avec *παραβολή* (lire « parabolé ») qui pouvait signifier « comparaison » ; récupéré par le latin sous la forme *parabola*, le mot évolua en « parawola » (les lèvres ne se ferment plus complètement, transformant ainsi le phonème /b/) puis *paraula* (lire « paraoula »). Ainsi, le sens ancien a-t-il davantage rapport à l'intention de désigner par le mot, de « comparer » l'objet visé c'est-à-dire littéralement de former une paire entre le mot et l'objet, qu'à la production orale même de ce mot. Ce sens ancien est volontiers poétique, puisque nous pouvons penser cette paire comme liant l'esprit à la matière.

2.2.2. La parole en poupée russe

Comme nous l'avons vu dans le précédent chapitre, en empruntant quelques concepts à Henri Laborit, la réalité peut se décrire par une succession de niveaux d'organisation, et l'*information-structure* regroupe l'ensemble des règles liant les éléments unitaires d'un niveau donné en une unité de niveau supérieur.

La parole n'échappe pas à cette règle : d'un point de vue acoustique, son unité de base, sa « particule élémentaire », est le *phone*. Il s'agit du son prononcé, tel qu'étudié par la phonétique, et varie en fonction de chaque langue, de chaque accent voire de chaque locuteur ; il est écrit par convention entre crochets (par exemple [r]). La phonologie ne s'intéresse quant à elle qu'aux variations phonétiques changeant le sens des mots à l'intérieur d'une même langue ; autrement dit, elle ne prend pas en compte les variations de prononciation sémantiquement insignifiantes. De ce point de vue, l'unité de base de la parole est le *phonème*, écrit par convention entre barres obliques (par exemple /r/). A l'intérieur d'une langue, un même phonème peut avoir plusieurs réalisations phonétiques, c'est-à-dire qu'il peut être prononcé en utilisant différents phones sans que le sens du mot n'en soit altéré. Concrètement, si nous demandions à un Parisien et à un Ardéchois de prononcer le mot *drôle* (/dʁol/), le premier utiliserait le phone [o] (o fermé, comme dans *mot*) pour réaliser le phonème /o/, et le second le phone [ɔ] (o ouvert, comme dans *parole*). Pourtant, l'usage d'un phone ou de l'autre ne change pas notre compréhension du mot ; il s'agit alors d'allophones (équivalents phonétiques des formes allotropiques de la chimie ; en d'autres mots, les allophones sont à un phonème ce que le graphite et le diamant, éléments allotropiques, sont au carbone).

Les phonèmes sont ensuite assemblés entre eux pour former l'unité du niveau supérieur qu'est la syllabe (par exemple, l'association des phonèmes /b/ et /a/ forme la syllabe /ba/). Puis les syllabes se juxtaposent pour former les mots (par exemple /ba/ et /to/ pour le mot *bateau* ; notons là de façon légèrement digressive que le « langage texto » associé aux jeunes et tant décrié par certains est plus proche de la phonologie que le sont les graphèmes usuels). Enfin, les mots s'organisent en phrases grâce à l'*information-structure* que constituent grammaire et syntaxe, et les phrases en paragraphes, les paragraphes en discours, *et cetera* jusqu'à atteindre ultimement l'ensemble des paroles d'un individu au cours de sa vie.

2.2.3. Composition des phonèmes

En repartant de la base et dans la direction opposée, les phonèmes sont réalisés par des phones eux-mêmes constitués de sons complexes, contenant de l'énergie à de nombreuses fréquences (fondamentale et harmoniques) que l'on pourrait enfin considérer comme des sons purs. Les phonèmes regroupent les sons pertinents pour la compréhension d'une langue, et recouvrent les consonnes, les voyelles et les pseudo-voyelles. La langue française contient notamment 36 phonèmes (17 consonnes, 16 voyelles et 3 semi-voyelles), comparativement aux 26 lettres de notre alphabet. Le contenu fréquentiel de ces sons de parole s'observe idéalement avec un spectrogramme (figure représentant l'énergie acoustique en fonction des fréquences et du temps ; Figure 2.3).

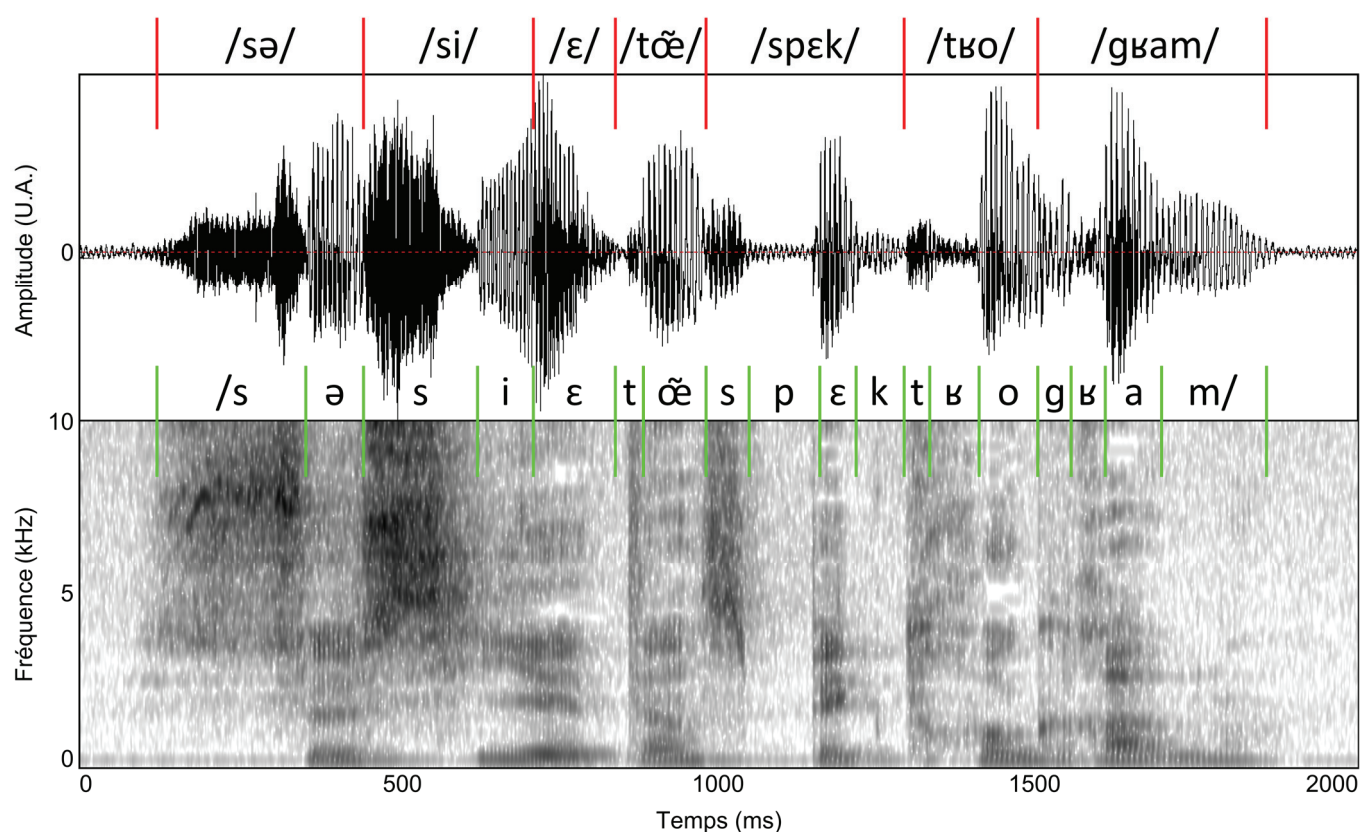


Figure 2.3. Forme d'onde (en haut) et spectrogramme (en bas) de l'enregistrement de la phrase « Ceci est un spectrogramme ». Les syllabes sont délimitées par les barres verticales rouges, et les phonèmes par les barres verticales vertes.

En observant la Figure 2.3., un premier constat est la dynamique temporelle des fréquences composant la parole : les syllabes durent quelques centaines de millisecondes (ce qui correspond à un rythme de 3 à 8 syllabes par seconde, ou Hertz - nombre d'évènements par seconde ; barres verticales rouges), avec des consonnes comportant des transitions fréquentielles de quelques dizaines de millisecondes (par exemple, pour le *s* de « spectro », bande foncée descendante aux alentours de 4,5 kHz), et des voyelles montrant une stabilité fréquentielle de l'ordre de la centaine de millisecondes (*e* de « ceci », bandes foncées stables entre 0 et 4 kHz). Un second constat s'opère suivant l'axe fréquentiel, avec l'observation de bandes compactes d'énergie à certaines fréquences (même exemple du *e* de « ceci », en dessous de 4,5 kHz), de bandes plus diffuses à d'autres fréquences (deuxième *s* de « ceci », énergie diffuse entre 4 et 10 kHz), ainsi que de vides énergétiques (*o* de « spectro » autour de 5 kHz). Ces bandes sont la manifestation des formants, reflets acoustiques de la configuration de l'appareil phonatoire et responsables de la perception de phonèmes, que nous verrons plus en détail accompagnés de leurs corrélats physiologiques au point 2.3.1, en étudiant la production de parole.

2.2.4. Information paralinguistique

La parole naturelle s'incarne dans la voix humaine qui, plus ou moins efficacement, véhicule bien plus d'*information circulante* que le simple message linguistique : genre, âge, identité, état émotionnel, voire même orientation sexuelle, classe sociale et niveau intellectuel. La prise en compte de cette information paralinguistique est rendue possible par la présence d'indices vocaux tels que la fréquence fondamentale (hauteur de voix), le timbre (texture), la prosodie (mélodie) et le rythme (vitesse des mots). Pour être complet, notons que la voix est perçue non pas en tant que telle, mais dans un flux sonore plus général qui contient également des informations sur l'environnement, comme par exemple la position du locuteur, ou le type (intérieur ou extérieur) et la taille de l'environnement en fonction de la réverbération.

Si les phonèmes, syllabes et mots peuvent relativement bien être délimités, d'où leur dénomination d'information segmentale, certains indices paralinguistiques sont d'ordre suprasegmental. C'est le cas en français de la prosodie, ou mélodie de la voix, qui véhicule l'intonation tout au cours d'une phrase pour donner de l'information sur l'état émotionnel du locuteur. Cet indice est analysé sur une échelle de temps de quelques secondes, donc recouvrant plusieurs segments phonémiques, syllabiques ou lexicaux.

Une information importante concerne le cas des langues tonales, où un même mot prononcé avec différentes trajectoires tonales, sortes de prosodies à l'échelle lexicale, aura des sens différents. Cela n'exclut cependant pas l'existence d'une prosodie suprasegmentale, indiquant toujours l'humeur du locuteur.

Maintenant que nous en connaissons un peu plus sur ce qu'est la parole dans sa structure macro- et microscopique, sa place dans le monde et son utilité chez l'Homme, il est temps d'aborder les différentes facultés et autres notions la sous-tendant.

2.3. Le cycle de la parole

Les organismes unicellulaires possèdent une information-structure plus simple que celle des grands primates, ce qui ne signifie pas qu'ils soient moins adaptés à leur environnement. A titre d'exemple, le corps humain est composé de dizaines de milliers de milliards (environ 10^{13}) de cellules, et héberge dix fois plus (environ 10^{14}) de bactéries. Au fur et à mesure de la

complexification des espèces, des groupes de cellules se spécialisent dans une fonction en formant les organes, et le système nerveux (possédé par tous les animaux au-delà de l'éponge) permet de connecter ceux-ci entre eux et à l'environnement, en supplément de la communication chimique. Des 302 neurones du ver nématode *Caenorhabditis elegans* aux plusieurs centaines de milliers de neurones de la double chaîne ganglionnaire des insectes, jusqu'aux cent milliards de neurones du cerveau humain, le système nerveux varie grandement en structure et en taille dans le règne animal, mais conserve toujours le même principe : recueil d'*information circulante* dans l'environnement extérieur ou intérieur à l'organisme *via* des organes perceptifs, traitement plus ou moins sophistiqué de cette information par le système nerveux, et envoi de commandes vers différentes parties de l'organisme pour l'adapter à ces environnements. C'est là ce que les Neurosciences appellent communément la *boucle perception-action*. Nos perceptions dirigent nos actions, qui produisent alors de nouvelles perceptions, et ainsi de suite.

La parole peut être vue comme une faculté reposant sur ce cycle : lorsqu'une personne nous parle, nous percevons sa voix dans un flux sonore plus global, puis analysons cette scène auditive pour en extraire le message linguistique entre autres informations. En fonction du résultat de cette analyse, nous pouvons produire une action en mettant notre corps en mouvement, pour par exemple, produire de la parole ou se déplacer, action qui sera perçue par l'interlocuteur chez qui se déroulera la même suite d'événements. Le processus qui vient d'être décrit fait intervenir et se croiser les boucles perception-action de deux interlocuteurs, formant en quelque sorte une trajectoire en forme de huit. Nous avons choisi d'appeler cette double boucle le *cycle de la parole* qui, en se rappelant de l'impact de la parole sur l'espèce humaine tel qu'exposé dans le point précédent, peut être considérée comme motrice pour l'humanité.

Les parties suivantes de ce chapitre seront consacrées à exposer les différentes étapes de ce cycle et leurs processus biologiques sous-jacents, en commençant par la production pour terminer par la perception. Ce choix en opposition de phase avec la boucle classique est justifié par la nature des notions nécessaires à la compréhension des études exposées dans le présent manuscrit : étant centrées sur la Speech ABR, marqueur électrophysiologique de l'encodage des sons de parole dans le système nerveux auditif, nous avons jugé plus pertinent d'exposer d'abord la production des sons de parole pour ensuite mieux comprendre les spécificités de leur encodage neuronal. Aussi, la partie du cycle située après la perception et avant la production n'a pas été abordée plus avant dans ces travaux de thèse.

2.3.1. Production

Ainsi commençons-nous notre immersion dans le cycle de la parole par l'étude de sa production. La faculté à produire le langage articulé humain repose sur l'activité coordonnée d'un ensemble d'organes regroupés sous le nom d'*appareil phonatoire*. Celui-ci se subdivise en trois parties : respiratoire (diaphragme, poumons et trachée), vibratoire (larynx, glotte et plis vocaux) et articuloire (cavités orale et nasales, mâchoire, lèvres, dents, langue et voile du palais). L'appareil respiratoire expulse un flux d'air à une pression suffisante pour faire entrer en vibration les plis vocaux (membranes constituées de muscles et de ligaments d'environ 25 millimètres de long, aussi appelés cordes vocales par abus de langage) au niveau du larynx. Le son alors produit va cheminer à travers les cavités orales et nasales, où la position des différents articulateurs va en modifier le contenu spectral (adjectif décrivant les fréquences de l'onde acoustique).

2.3.1.1. La théorie source-filtre

Le processus de production qui vient d'être décrit a été conceptualisé en 1960 par l'ingénieur et phonéticien suédois Gunnar Fant, sous le nom de *théorie source-filtre* (Fant, 1960). La *source* est située au niveau de la glotte (Figure 2.4., niveau *violet* à gauche), segment du larynx (lui-même situé à l'extrémité supérieure de la trachée) comportant les plis vocaux, où le passage du flux d'air expulsé par les poumons (niveau *orange*) sous pression du diaphragme fait entrer en vibration ces mêmes plis vocaux. Une onde sonore¹ est alors générée, de *fréquence fondamentale* (F_0 ; principal corrélat de la hauteur perçue, ou *pitch*) égale au nombre de cycles ouverture-fermeture des plis vocaux mis en vibration, et contenant des *harmoniques* (multiples entiers de la fréquence fondamentale ; $2F_0$, $3F_0$, $4F_0$...) d'amplitude inversement proportionnelle à la fréquence (soit allant décroissante en montant dans les harmoniques ; niveau *violet* à droite). La fréquence de vibration des plis vocaux peut être modifiée par mouvement des cartilages aryténoïdes sur lesquels ils sont fixés, et qui permet de réguler leur tension avec la même logique que pour une corde de guitare : plus les plis vocaux sont tendus, plus la fréquence fondamentale sera élevée, et donc plus le son sera perçu comme aigu.

Le filtrage du signal source est assuré par l'appareil articuloire, ensemble des cavités orales et nasales jouant le rôle de caisses de résonance, et d'organes mobiles (mâchoire pour les dents, lèvres, pointe et base de la langue, et voile du palais) sous contrôle moteur plus ou moins volontaire (niveau *bleu* à gauche). En fonction de la configuration de ces différents

¹ Le son est une vibration mécanique d'un fluide compressible et élastique (air, eau ou autre liquide), s'y propageant sous forme d'onde par variations locales de pression (notons que les particules du fluide ne bougent pas, mais vibrent autour d'un point d'équilibre au passage de l'onde). Les sons se caractérisent par leur fréquence fondamentale (nombre de périodes par unité de temps ; en partie liée au percept de hauteur du son, « grave » ou « aigu ») et leur amplitude (valeur maximale de l'onde ; liée au percept de sonie, « faible » ou « fort »).

articulateurs, c'est-à-dire de leur position relative dans la dynamique des mouvements possibles, certaines bandes de fréquence vont être amplifiées, d'autres seront atténuées et encore d'autres inchangées (niveau *bleu* à droite). La configuration de l'appareil articulatoire laisse son empreinte dans le spectrogramme, sous la forme des *formants* (niveau *rouge*). Concrètement, ils s'observent dans le spectre (énergie en fonction de la fréquence, donc pour un échantillon temporel ou une moyenne d'échantillons donnés) par une recherche des harmoniques les plus amples, en utilisant par exemple une enveloppe de Hilbert, sorte de drap jeté sur les pics spectraux.

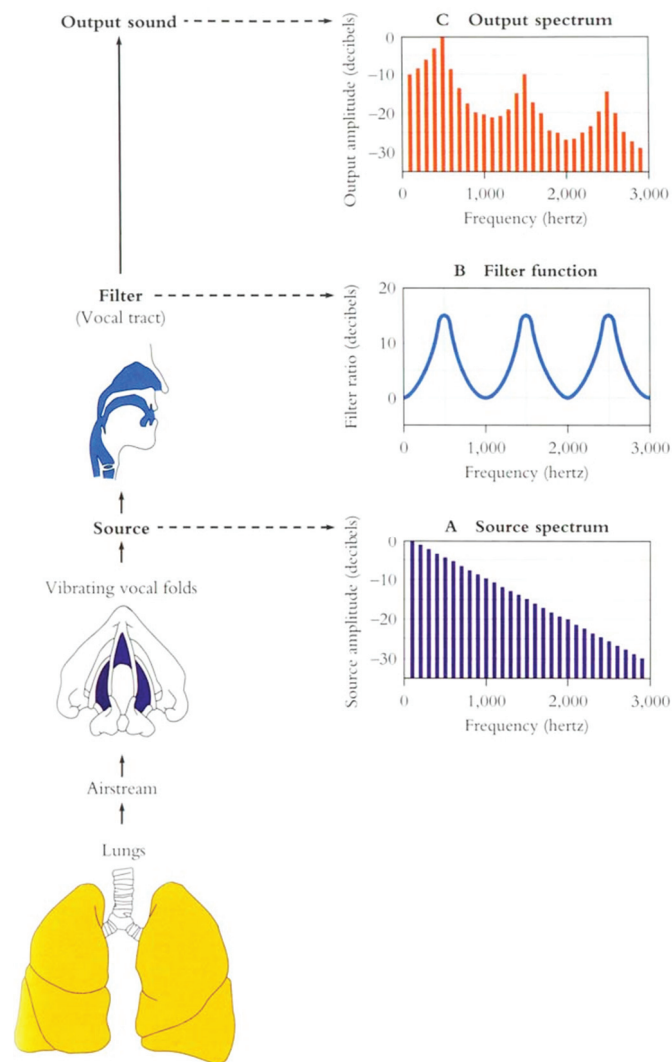


Figure 2.4. Illustration simplifiée de la théorie source-filtre (issue des cours de linguistique d'automne 2001 à l'Université de Pennsylvanie, dispensés par Mark Liberman ; disponible à l'adresse http://www.unibielefeld.de/lili/personen/vgramley/teaching/HTHS/acoustic_2010.html).

Notons que du point de vue évolutif, les organes impliqués dans la production de parole possédaient antérieurement et ont conservé parallèlement des fonctions tout autres : respiration

pour le diaphragme, les poumons, la trachée et les cavités orales et nasales ; alimentation pour les lèvres (surtout chez le nourrisson), les dents, la mâchoire, la langue, la cavité orale et le larynx (empêche les aliments d'entrer dans la trachée, en assurant cette fonction plus efficacement chez le nourrisson, où il est en position haute) ; olfaction avec les cavités nasales ; expression faciale avec les lèvres et les dents. La parole a ainsi bénéficié de la possibilité d'un contrôle moteur fin et coordonné de la plupart de ces organes, initialement engagés dans d'autres fonctions.

En somme, le signal de parole peut être vu comme une vague de fond (modulation de la fréquence fondamentale au niveau de la glotte) porteuse de vaguelettes de surface, de détails fins (empreinte de la configuration de l'appareil phonatoire). Dans la suite de cette image, une dichotomie du spectre aux alentours de 500 Hz séparerait grossièrement la fréquence fondamentale (voisement, basse fréquence, porteuse) des formants (identité des phonèmes, haute fréquence, détails). Une nomenclature parfois utilisée nomme ces deux parties *périodicité* (inférieure à 500 Hz) et *structure fine temporelle* (supérieure à 500 Hz ; Ritsma, 1962 ; Moon & Hong, 2014). Une troisième partie, l'*enveloppe temporelle*, concerne les modulations d'amplitude liées au rythme phonémique ou syllabique (inférieures à 50 Hz).

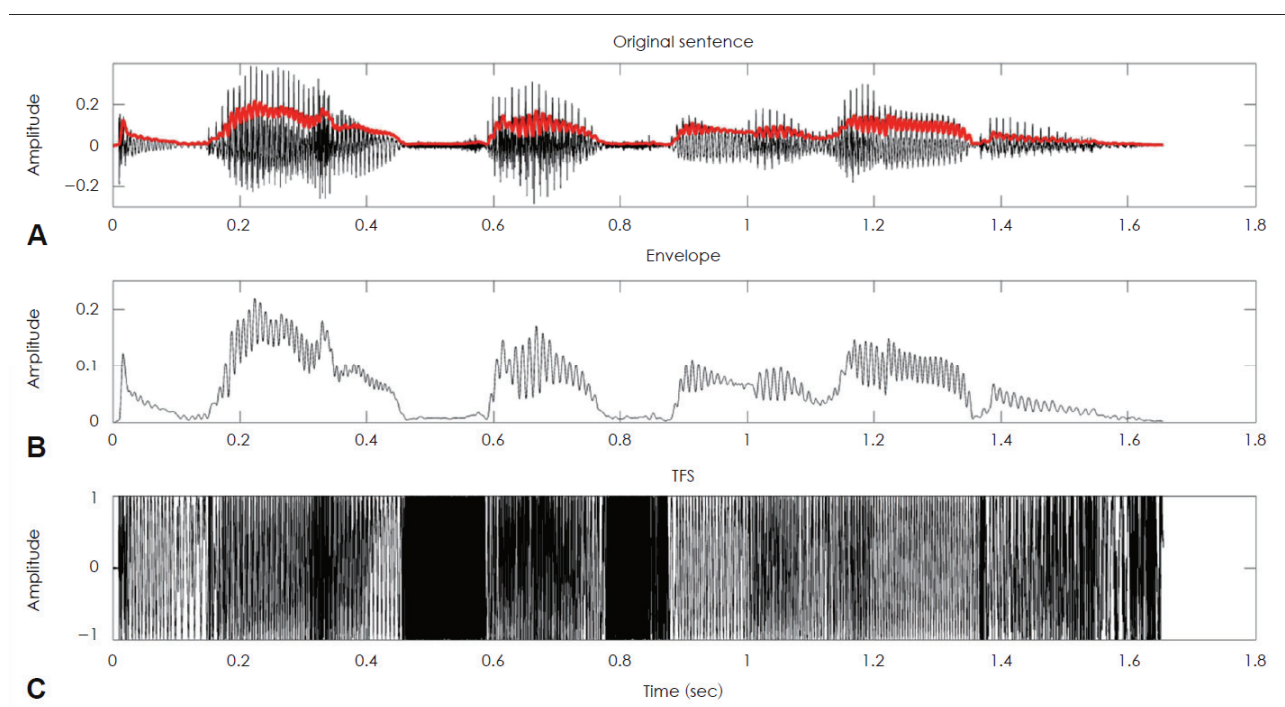


Figure 2.5. Informations d'enveloppe et de structure fine temporelle telles qu'extraites d'une phrase (adapté de Moon & Hong, 2014). Dans le repère A, l'enveloppe temporelle est représentée en rouge sur le signal acoustique brut. Elle est obtenue par transformée de Hilbert puis filtrage passe-bas à 200 Hz, et est représentée dans le repère B. La fréquence de coupure à 200 Hz laisse passer une partie de l'information de périodicité, ce qui se manifeste par les vaguelettes surmontant l'enveloppe entre 200 et 300 ms, par exemple. Le repère C représente la structure fine temporelle, contenant dans ce cas les informations spectrales au-delà de 200 Hz. Elle porte ainsi à elle seule la quasi-totalité des formants, et joue donc un rôle prépondérant dans la perception des phonèmes. Enfin, la modulation de la structure fine temporelle par l'enveloppe temporelle permet de retrouver le signal brut.

2.3.1.2. Boucle audio-phonatoire

La schématisation du cycle de la parole en une double boucle perception-action croisée entre deux interlocuteurs est, bien que déjà compliquée, encore loin de la réalité. Une première greffe apportée au schéma est un système de rétrocontrôle (*feedback*) de la production de parole, dénommé *contrôle audio-phonatoire* ou *boucle audio-phonatoire* (Baddeley *et al.*, 1998). Les mouvements de l'appareil phonatoire, sous contrôle moteur cortical, sont sans cesse ajustés en se basant principalement sur des mesures de deux types : somesthésiques (sensations corporelles liées à la position des différents articulateurs) et auditives. C'est pour ce deuxième type que nous parlons de *boucle audio-phonatoire*. La parole produite par un locuteur est entendue par les voies auditives de ce même locuteur, et l'information ainsi apportée permet de réguler la production : si par exemple notre voix est trop aigüe, nous détendrons les plis vocaux au niveau de la glotte ; si l'on ne comprend pas bien ce que nous tentons de prononcer, nous accentuerons notre articulation. Ce système est fortement sollicité dans l'acquisition du langage chez l'enfant, ou d'une seconde langue.

Pour replonger succinctement dans la *Nouvelle grille* d'Henri Laborit, nous retrouvons en cette boucle un exemple de *servomécanisme*, soit de dispositif assurant la concordance entre une intention et sa réalisation à moindre coût énergétique. Ainsi, nous n'avons pas besoin de contrôler avec une extrême précision les nombreux organes de notre appareil phonatoire, mais envoyons un programme de commandes premières, qui sera ensuite affiné par comparaison entre le son perçu et le résultat désiré. A titre de comparaison et pour mieux comprendre cette notion d'effort, imaginons que nous devons écrire un long document. Soit nous pesons précautionneusement chaque mot, chaque phrase et chaque idée avant de la coucher sur le papier, ce qui peut s'avérer très chronophage et intellectuellement épuisant ; soit nous déroulons globalement nos pensées sans grande attention portée à la forme, en premier lieu, avant de revenir plus tard sur celles-ci pour les ajuster et ainsi obtenir un résultat plus conforme au rendu souhaité. En considérant cette même logique, la *boucle audio-phonatoire* pourrait offrir au locuteur une économie d'énergie.

2.3.1.3. Phonétique articulatoire

En phonétique, plusieurs approches ont tenté de caractériser les sons de parole produits par l'appareil phonatoire. La phonétique articulatoire s'intéresse à l'influence du mouvement des différents articulateurs sur le son produit. Les phonèmes y sont recensés selon trois paramètres : leur *lieu d'articulation*, leur *mode d'articulation*, et la présence ou non du *voisement* (Tableau 2.1.). Le lieu d'articulation représente l'endroit de l'appareil phonatoire où se produit l'obstruction totale ou partielle du flux d'air. Par exemple, le phonème /b/ est articulé au niveau des lèvres, qui sont d'abord fermées pendant le voisement (vibration des plis vocaux), puis s'ouvrent pour relâcher l'air accumulé dans la cavité orale.

	Bilabiale	Labio-dentale	Labio-palatale	Labio-vélaire	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire
Occlusive	p b				t d				k g	
Nasale	m				n			ɲ	ŋ	
Fricative		f v				s z	ʃ ʒ			ʁ
Spirante			ɥ	w				j		
Latérale						l				

Tableau 2.1. Représentation des 17 consonnes (cases blanches) et des 3 semi-voyelles (cases bleues) du français, accompagnées du phonème ɲ emprunté surtout à l'anglais (par exemple, suffixe « -ing » de *timing* ; case verte). Les colonnes jaunes renseignent du lieu d'articulation, et les lignes orange du mode. Le voisement est indiqué par la position à l'intérieur de la case : gauche ou droite pour non-voisé ou voisé, respectivement.

Le mode d'articulation indique le mécanisme global de production du phonème. En gardant l'exemple du /b/, le mode d'articulation est occlusif oral voisé, car le flux acoustique passe dans la cavité orale où il est d'abord bloqué, puis relâché, avec tout du long la présence d'un voisement. Nous pouvons donc décrire de façon encore plus complète le phonème /b/, comme étant une consonne bilabiale occlusive orale voisée. En comparaison, le phonème /d/ est une consonne dentale-alvéolaire occlusive orale voisée : seul le lieu d'articulation change par rapport au /b/, avec une occlusion du flux acoustique s'opérant par contact entre la pointe de la langue (*apex*) et les alvéoles des incisives supérieures (emplacements dans la gencive où se logent les dents). Prenons aussi l'exemple du /m/, qui nous ressortira lors de la partie expérimentale de cette présente thèse : il s'agit d'une consonne bilabiale occlusive nasale voisée, pour la prononciation de laquelle le voile du palais (*velum*) s'abaisse, dérivant alors le signal source vers les cavités nasales. C'est parce /b/ et /m/ ne diffèrent qu'au niveau du mode d'articulation, respectivement oral et nasal, que nous prononçons les /m/ comme des [b] lorsque nous sommes enrhumés (« Tu peux *be* passer un *bouchoir* s'il te plaît ? ») : la cavité nasale étant bouchée, la déviation du flux d'air suite à l'abaissement du *velum* ne produit plus pleinement son effet, et la consonne nasale sonne comme une consonne orale.

Plus généralement, certaines valeurs de ces paramètres articulatoires ont été liées aux formants, ce qui permet notamment une description intuitive des voyelles sur un repère F1/F2 (valeur du premier formant en fonction de la valeur du deuxième formant). La fréquence du premier formant (F1), plus basse que celles des formants suivants, est corrélée positivement au degré d'ouverture de la bouche et à la position verticale de la langue (*degré d'aperture*), en étant à son minimum lorsque les lèvres et la mâchoire sont fermées et que la langue est en position basse. La fréquence de F2 est influencée par la position horizontale de la langue (avant ou arrière, par exemple /i/ et /u/), assimilable au lieu d'articulation, ainsi que par la position des lèvres (étirées ou arrondies, même exemple /i/ et /u/). Indiquons au passage que F3 est

uniquement lié à la position des lèvres, et que F1 et F2 sont dans la plupart des cas suffisants pour distinguer les différentes voyelles. Ainsi, sur le repère F1/F2, les voyelles du français dessinent un *triangle vocalique*, avec aux extrémités /a/, /i/ et /u/ (« ou ») ; pour les voyelles anglaises, cette représentation prend la forme d'un trapèze (Figure 2.6.).

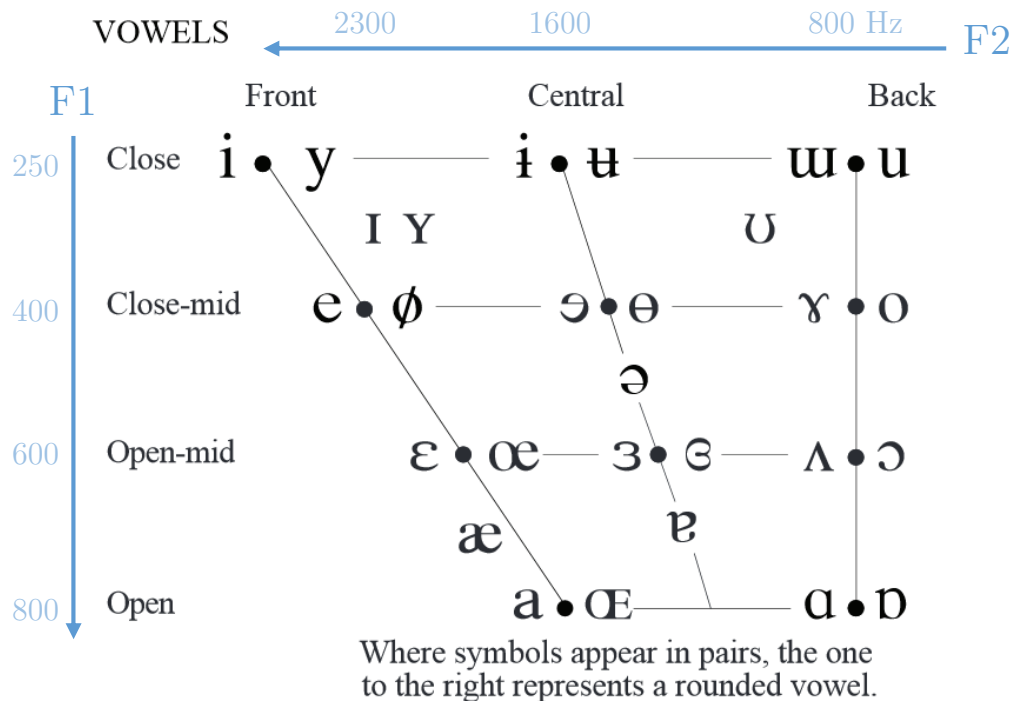


Figure 2.6. Trapèze vocalique des voyelles de l'anglais (adapté de *The International Phonetic Alphabet*, 2005).

Dans le cas des phonèmes nasalisés, la transmission de l'onde acoustique dans l'air des cavités nasales, consécutive à l'abaissement du *velum*, entraîne la présence de nouveaux formants propres à la résonance de ces cavités, et des lacunes dans les formants propres à la résonance de la cavité buccale, que l'on désigne sous le nom de *formants d'anti-résonance* ou *pôle-zéro*.

La phonétique acoustique utilise une autre grille pour décrire les phonèmes, en se basant cette fois-ci sur leurs propriétés spectrales plutôt que sur les mouvements les ayant générés. Elle consiste donc plus en une description du signal acoustique qu'en une étude de la physiologie de l'appareil phonatoire, et utilise la notion de *traits distinctifs* comme propriétés acoustiques permettant de distinguer deux phonèmes. La liste de ces traits s'est allongée et complexifiée au fil du XXème siècle, grâce aux travaux successifs et croisés des linguistes russes Nikolai Troubetzkoï et Roman Jakobson, du suédois Gunnar Fant, du letton-américain Morris Halle et de l'américain Noam Chomsky (constituant de fait un bel exemple d'intelligence dans une période de l'Histoire plutôt froide ; Troubetzkoy, 1939 ; Jakobson *et al.*, 1951 ; Jakobson & Halle 1956 ; Chomsky & Halle, 1968).

Enfin, la phonétique auditive vise à comprendre la façon dont l'oreille et le système nerveux auditif perçoivent les sons, en étudiant les indices acoustiques effectivement utilisés par nos voies auditives pour comprendre la parole.

2.3.1.4. Parole naturelle ou synthétique

Dès lors que notre compréhension de la structure de la parole s'est étoffée, il est devenu concevable de la produire de façon synthétique, plus ou moins grossièrement. En effet, une simple description des deux premiers formants permet de distinguer la plupart des voyelles, mais le rendu est artificiel. Des ingrédients acoustiques peuvent être ajoutés à cette description basique pour s'approcher le plus possible du rendu naturel. En plus des nombreuses applications industrielles dans l'informatique, la robotique et l'intelligence artificielle, la parole artificielle revêt un intérêt primordial pour les sciences touchant au langage, en cela qu'elle permet de contrôler très finement les différents paramètres temporels et fréquentiels pour les faire varier individuellement. Les observations faites en l'utilisant sont cependant limitées par la non-linéarité du système auditif : la réponse cérébrale à un son de parole synthétique dénué des indices de timbre sera par exemple différente de celle évoquée par la même syllabe naturellement produite (Theunissen & Elie, 2014).

La zone du spectre où réside la majorité des indices pertinents pour la reconnaissance des phonèmes s'étale entre 250 et 4000 Hz (ou de 100 à 4000 Hz pour les langues tonales, utilisant F0 comme indice phonémique). De manière indicative, le premier formant varie entre 250 et 800 Hz (respectivement pour le /i/ et pour le /a/), le deuxième entre 700 et 2500 Hz, le troisième entre 1500 et 3500 Hz, et le quatrième entre 2500 et 5000 Hz. Au-delà de 5000 Hz, l'énergie spectrale informe sur l'identité du locuteur (en participant à la perception du *timbre*), et donne une brillance au son qui lui confère un aspect naturel. En dessous de 500 Hz, la fréquence fondamentale informe sur le genre, l'identité et l'état émotionnel du locuteur (prosodie) ; aussi, F0 peut varier entre 80 et 200 Hz pour une voix d'homme, entre 170 et 300 Hz pour une voix de femme, et entre 300 et 500 Hz pour une voix d'enfant.

Du fait de la physiologie de production des sons de parole, la fréquence fondamentale tout comme les formants ne sont pas parfaitement stables dans le temps : il existe des micro-mélodies, ou variations très fines autour d'un point fréquentiel d'équilibre, dues à l'irrégularité du flux d'air expulsé par les poumons et aux mouvements minimes des différents organes de l'appareil phonatoire.

L'ajout des très hautes fréquences (supérieures à 5 kHz) et des micro-mélodies, bien que non-nécessaires à la reconnaissance des phonèmes, améliorera la naturalité des sons de parole de synthèse.

Aussi, la présence de coarticulations dans la parole naturelle complique significativement la tâche de synthèse. En effet, les différentes composantes spectrales nécessaires à l'identification d'un même phonème varient en fonction des phonèmes qui l'entourent : il n'y a donc pas de recette unique pour synthétiser un phonème, mais plutôt des indications à appliquer au contexte. Concrètement, lorsque nous parlons, il est rare que nous détachions proprement les phonèmes les uns des autres (« je ne sais pas » deviendra « je sais pas » puis « j'sais pas » et enfin « chaipa »). Par souci de rapidité et d'économie d'énergie, nous entamons l'articulation d'un phonème pendant celle du phonème précédent. Par exemple, les mots *rab* (/rab/) et *rame* (/ram/) ne diffèrent qu'au niveau de leur troisième phonème, respectivement oral et nasal ; en prononçant ces mots anormalement lentement, la voyelle /a/ reste identique, mais à vitesse réelle, le phonème /a/ est nasalisé dans le second cas du fait de la baisse du voile du palais en anticipation de la prononciation du /m/. En quelque sorte, les phonèmes se mélangent de proche en proche. L'application d'une recette stricte de synthèse du phonème /a/ pour synthétiser le mot *rame* pourra alors donner un résultat s'éloignant du naturel.

2.3.1.5. Techniques de synthèse

Plusieurs stratégies existent pour générer un signal de parole sans intervention humaine directe, et ont été explorées à partir des années 1940 marquant l'arrivée du psychologue américain Alvin Liberman au très célèbre laboratoire Haskins, avec pour but premier l'assistance aux vétérans de la Seconde Guerre Mondiale par l'invention d'une machine de lecture automatique de textes (Shankweiler & Fowler, 2015).

La première approche a visé à décrire le spectrogramme des phonèmes à synthétiser, et est généralement nommée *synthèse formantique*. L'axe temporel est discrétisé en un certain nombre de fenêtres temporelles (d'une largeur de 5 millisecondes par exemple), pour chacune desquelles sont décrites les fréquences, amplitudes et largeurs de bande de F0 et des formants. Cette technique de synthèse ne permet pas une description très fine du spectre, telle que la définition de vides énergétiques (formants d'anti-résonance de la nasalité, par exemple), mais suffit à rendre distinguables les différents phonèmes de façon robuste. Relativement simple d'usage et implémentée dans des logiciels faciles à prendre en main, c'est cette technique que nous avons choisie pour la synthèse des *stimuli* de l'étude chez le malentendant appareillé.

Avec le développement rapide de l'informatique rendant possible le stockage et la manipulation d'un grand nombre d'échantillons de syllabes naturelles, une seconde approche a consisté en la juxtaposition de celles-ci pour reconstituer des mots et des phrases en une sorte de « patchwork syllabique », avec un rendu plus naturel que celui des synthétiseurs formantiques. Aussi, des algorithmes d'analyse phonologique de texte (*text-to-speech*) ont peu

à peu permis l'association automatique entre graphèmes (par exemple « eau », « au », « ot » et « o ») et phonèmes (/o/, dans notre exemple).

L'approche la plus moderne a implémenté les connaissances articulatoires apportées en grande partie par la théorie motrice de la parole (Liberman, 1954 ; Liberman & Mattingly, 1985 ; Liberman, 1996) pour synthétiser les sons de parole sur la base de la description des mouvements physiologiques liés. Cette *synthèse articulatoire* permet de prendre en compte les phénomènes de coarticulation et de micro-mélodies bien plus aisément que dans la synthèse formantique classique, et est moins limitée en nombre de formants descriptibles. Son utilisation est par contre bien plus complexe, et permet moins de contrôler finement les composantes spectrales des sons synthétisés. En d'autres termes, l'utilisateur de ce type de synthétiseurs ne décrit plus les sons au niveau du spectrogramme mais au niveau de l'appareil phonatoire, et ses commandes sont traduites en fréquences par des modèles souvent difficiles d'accès.

La parole, en tant que signal acoustique véhiculant un message linguistique parmi d'autres informations, possède une structure complexe, étudiée au moyen des grilles de lecture de la phonétique articulatoire, acoustique ou auditive. Les mécanismes de sa production physiologique sont néanmoins compris en grande partie, au moins suffisamment pour réussir à implémenter sa synthèse dans de nombreuses applications technologiques (serveurs téléphoniques, assistants vocaux sur smartphone, logiciels de lecture automatique de textes, robotique, entre autres). Cependant, si le signal en sortie de la bouche du locuteur ou du haut-parleur du synthétiseur est relativement pur, il va potentiellement se mêler à d'autres sons et subir diverses transformations lors de son voyage jusqu'aux oreilles de l'auditeur. C'est ce dont nous allons traiter à présent.

2.3.2. Transmission

2.3.2.1. Notion de bruit

Reprenons la définition du son : il s'agit de la transmission de proche en proche de la vibration des particules d'un fluide élastique. Imaginons que nous lançons une pierre dans un lac : des cercles concentriques, centrés sur le point d'impact de la pierre sur l'eau, seront observables qui donneront l'impression de s'agrandir et de s'éloigner du centre. En réalité, les molécules d'eau ne se déplacent pas horizontalement, mais « vibrent » verticalement autour d'un point d'équilibre. Au bout d'un certain temps, l'amplitude de ces vibrations diminuera jusqu'à ce que la surface du lac redevienne plane. Le voisement, ou vibration des plis vocaux générée par le passage d'un flux d'air sous pression (le plus souvent expiré par les poumons, mais aussi parfois inspiré), va de manière comparable mettre en mouvement pendulaire les

molécules d'air, à ceci près que chaque cycle ouverture-fermeture de la glotte ressemblerait à une pierre tombant dans le lac.

En continuant avec cette image aquatique, imaginons que cette fois un rocher soit lancé à quelques mètres du point d'impact de la pierre : de grandes vagues seraient générées, sur le même principe que les petites vaguelettes concentriques précédentes, et troubleraient grandement l'observation de celles-ci. Cela illustre la notion de bruit, qui peut se définir comme l'ensemble des objets de même nature que le signal d'intérêt entraînant la dégradation de ce même signal d'intérêt. Le dernier mot de cette définition est primordial : la notion de bruit est relative. Si nous assistons à un concert, le signal d'intérêt sera vraisemblablement la musique jouée par les instrumentistes, et les toussotements d'un spectateur seront alors considérés comme du bruit, comme une pollution acoustique ; par contre, si nous essayons de nous endormir avec les fenêtres ouvertes pour lutter contre un été caniculaire, et qu'une voiture passe dans la rue en contrebas en jouant à fort volume le même morceau de musique, il sera cette fois-ci perçu comme du bruit. Ainsi, dans les paragraphes suivants, nous considérerons comme *bruit* toute production sonore venant altérer le signal de parole d'intérêt.

En plus de constituer un danger pour l'organe de l'audition lorsqu'il est de forte intensité, le bruit gêne la bonne perception du signal d'intérêt de façon non-linéaire : il n'est pas simplement juxtaposé au signal de parole, mais du fait qu'il partage le même fluide de transmission, peut modifier ce signal en fonction de sa phase relative. En reprenant l'image du lac, imaginons que nous lancions cette fois deux pierres identiques en masse et en forme, avec une trajectoire identique, à deux points distincts dans le lac : elles occasionneraient toutes deux les mêmes vibrations visualisables par des cercles concentriques s'éloignant du centre, avec les mêmes vitesses et amplitudes, si bien que la rencontre des deux ondes les annulerait lorsqu'en opposition de phase, et les amplifierait lorsque parfaitement en phase. D'ailleurs, cette logique est utilisée dans l'isolation acoustique active, où une onde sonore est atténuée voire anéantie en lui opposant la même onde en antiphasé. En résumé, le bruit ne masque pas seulement le signal d'intérêt : il peut le faire disparaître.

Les bruits peuvent être très divers : d'origine humaine ou non, transitoires ou continus, graves ou aigus, forts ou faibles. Quels qu'ils soient, il suffit qu'ils masquent un indice spectral indispensable à la compréhension d'un phonème d'intérêt pour rendre ce dernier inintelligible. Cette situation se produit constamment, car il est très rare qu'une conversation ne se déroule sans aucun bruit alentour. Pourtant, nous comprenons malgré cela le message d'intérêt. Comme nous le verrons dans le point suivant, ce paradoxe apparent est levé par la conception ingénieuse de notre système nerveux auditif, à la fois par l'implémentation de mécanismes de débruitage, et par l'existence de flux descendants d'information permettant de prédire la perception en fonction du contexte et des connaissances préalables. Du fait de la défaillance de l'un ou de l'autre de ces processus, voire même des deux, la compréhension de la parole

dans le bruit est la plainte la plus rapportée par les malentendants. Nous reviendrons également sur ce point dans les chapitres suivants.

D'autres mécanismes, cette fois-ci passifs, peuvent aussi détériorer le signal de parole d'intérêt. Par exemple, une forte réverbération dans l'environnement d'écoute peut brouiller le signal d'intérêt, en le mélangeant avec son propre écho, temporellement décalé. Aussi, certains milieux sont dits absorbants, dans le sens où l'énergie acoustique est fortement atténuée en rencontrant certains objets, avec un impact plus grand pour les hautes fréquences, d'amplitude plus faible et donc plus facilement arrêtées. Prenons l'exemple d'une boîte de nuit : à l'intérieur, nous entendons équitablement toutes les fréquences, et la musique nous semble équilibrée ; par contre, en se plaçant à l'extérieur avec les portes fermées, nous n'entendons plus que les basses fréquences (sauf bien entendu si l'isolation acoustique est totale). Physiquement, la matière stoppe plus facilement les hautes fréquences, de longueur d'onde plus faible, que les basses fréquences, souvent plus amples. Dans la même logique et en relation avec le chapitre précédent, des lèvres fermées filtreront préférentiellement les hautes fréquences, d'où le contenu basses-fréquences du voisement des occlusives.

2.3.2.2. Stratégies d'adaptation au bruit

A nouveau, prenons un temps pour complexifier (ou rendre plus réaliste) notre *cycle de la parole*. Nous avons vu dans la partie précédente qu'une *boucle audio-phonatoire* existe, et que celle-ci permet une régulation *en ligne* de la production. Notre cerveau possède un autre servomécanisme, en cela qu'il est également capable d'adapter la production de parole à l'environnement. Un illustre exemple a été découvert en 1909 par le premier oto-rhino-laryngologiste français, Etienne Lombard : lorsque nous nous exprimons dans un bruit ambiant intense, nous haussons le volume de notre voix. Effectivement, cela peut paraître absurde puisque nous contribuons alors à l'augmentation du volume ambiant... une véritable fuite en avant, à laquelle nous pouvons trouver une alternative en nous rapprochant par exemple de l'oreille de l'auditeur, ou en demandant à celui-ci de rendre ses oreilles plus directives en plaçant ses mains en forme de coupelles derrière les pavillons.

Aussi, lorsque nous désirons retrouver une personne à l'intérieur d'une foule bruyante, il n'est pas rare que nous l'appelions de son prénom à gorge déployée, en répétant régulièrement l'appel pour augmenter nos chances d'être entendu. Cette stratégie, application de la théorie mathématique de la communication qui consiste à répéter un signal pour augmenter la probabilité qu'il soit reçu dans un environnement bruyant, est également utilisée par le manchot royal pour retrouver sa famille dans une colonie de centaines, voire de milliers d'individus, aux cris desquels s'ajoute parfois un vent puissant (Lengagne *et al.*, 1999).

Enfin, un exemple étonnant est celui des langues sifflées, tel que le *silbo* des Îles Canaries. Transposition sifflée de l'espagnol, ce mode de communication utilise les variations tonales et des intervalles de silence et d'attaques sémantiquement pertinents, et permet notamment aux bergers de se transmettre des informations d'une vallée à l'autre, sur plusieurs centaines de mètres jusqu'à plus d'un kilomètre en l'absence de bruit et d'obstacle. Cela constitue un cas de plus d'adaptation de la production du langage oral aux paramètres de transmission, ici la distance, en abandonnant la parole, inintelligible au-delà d'une centaine de mètres du fait de sa complexité structurelle et particulièrement de ses formants de faibles intensités, pour lui préférer le sifflement, fréquence quasi pure et puissante.

Pour résumer ce chapitre, disons que la parole n'arrive presque jamais seule aux oreilles de l'auditeur, mais presque toujours mêlée à d'autres sons, altérée par l'environnement de transmission, dégradée par des ondes en antiphase, masquée à certaines fréquences. Le corollaire d'un tel constat est que le système auditif humain hérite d'une dure tâche : décortiquer le flux sonore entrant pour isoler la parole du « bruit ». Aussi, nous avons vu que l'environnement influait sur la production de parole, complexifiant ainsi notre *cycle* qui va désormais se poursuivre au travers du système auditif humain, dans un chapitre exposant certaines notions primordiales pour la compréhension des études reproduites plus loin.

2.3.3. Perception

2.3.3.1 Modalités sensorielles

Tout évènement, toute action du monde physique peut être associée, si ce n'est confondue, à l'*information circulante* telle que définie par Henri Laborit. Nos organes sensoriels sont chacun spécialisés dans la perception de cette information sur différentes dimensions physiques, qu'ils convertissent en information utilisable par le système nerveux. L'information en question peut être classée en plusieurs catégories : électromagnétique, mécanique, chimique ou thermique.

La vision humaine permet de percevoir une partie du spectre électromagnétique, avec des longueurs d'onde s'étalant de 400 à 700 nanomètres, respectivement associées aux percepts de couleur violet et rouge. En dessous de 400 nm, nous sommes dans l'incapacité biologique de percevoir les ultraviolets, les rayons X et gamma, tandis qu'au-dessus de 700 nm nous ratons les infrarouges, les micro-ondes et les ondes radio. Les champs électriques et magnétiques échappent à notre conscience, bien que certaines personnes aient rapporté une phénoménologie particulière lorsque situées à proximité de lignes haute-tension ou d'antennes-relais de téléphonie mobile. Des conséquences indirectes des champs électromagnétiques sont

cependant perceptibles par d'autres sens ; par exemple, la somesthésie (plus communément assimilée au sens du *toucher*) permet de sentir nos poils se hérissier à proximité d'un champ électrique suffisamment fort ; encore, l'audition permet de percevoir les sons générés par les décharges électriques, de façon évidente dans le cas d'éclairs. Le monde animal non-humain recense certaines espèces dotées de récepteurs pour ces champs : l'ornithorynque est capable de percevoir le champ électrique produit par l'activité musculaire des animaux environnants, grâce à des électro-récepteurs situés à l'arrière de son bec (Manger *et al.*, 1995 ; Pettigrew *et al.*, 1998) ; aussi, nombre d'oiseaux utilisent le champ magnétique terrestre pour se guider dans leurs migrations (Larkin & Sutherland, 1977 ; Leask, 1977), et les ondes électromagnétiques ont été incriminées, en plus des pesticides, dans le dramatique *syndrome d'effondrement des colonies d'abeilles* (*Colony Collapse Disorder*), peut-être plus dans ce dernier cas par perturbation de leur système nerveux central, comparable à l'électro-sensibilité rapportée par certains de nos semblables.

L'information mécanique peut être captée par deux sens, nommément la *somesthésie* (regroupant *proprioception* ou sens de la position des différentes parties du corps, *nociception* ou sens de la douleur, et *thermoception* ou sens de la température, entre autres) et l'*audition*, étymologiquement l'*action d'entendre*. De la même façon que la vision ne permet de percevoir qu'une partie de l'information lumineuse électromagnétique existante, l'oreille dans la force de l'âge jeune n'entend que les fréquences acoustiques comprises entre 20 et 20000 Hz ; en-dessous sont définis les infrasons, au-dessus les ultrasons. La sensibilité de l'oreille est difficilement imaginable : elle peut détecter une pression de 20 μPa (micropascal), soit un déplacement d'air d'un centième de milliardième de mètre (10^{-11} m), de l'ordre de la taille des particules et des atomes.

Les fréquences hors du spectre auditif humain sont perçues et produites par différentes espèces animales. Si les baleines utilisent par exemple les infrasons pour communiquer sur de très longues distances, les tigres s'en servent aussi pour paralyser leur proie, ce qui produit également son effet sur les humains (Muggenthaler, 2000). En effet, de nombreuses études ont exploré les effets néfastes de ces infrasons, notamment dans un but militaire durant la Seconde Guerre Mondiale, et ont mis en évidence un lien de causalité entre le 7 Hz et la manifestation de malaises, de nausées et d'autres symptômes déplaisants chez l'auditeur. Bien qu'inaudibles, ces fréquences sont à même de donner une sensation de présence, et des mesures acoustiques à l'intérieur de maisons « hantées » ont parfois mené à la découverte d'infrasons amplifiés par le système d'aération. Sur l'autre versant du spectre auditif, les ultrasons ont de nombreuses applications technologiques, y compris médicales avec l'échographie et le traitement de certains cancers. Certains animaux perçoivent les ultrasons et s'en servent pour leurs vocalises, comme la souris jusqu'à 110 kHz ou le rat jusqu'à 60 kHz. Aussi, la borne supérieure du spectre auditif humain tend à baisser avec l'âge : après 25 ans, la plupart des individus perçoivent beaucoup

plus difficilement les fréquences au-delà de 16 kHz. C'est en se basant sur ce fait et dans la droite lignée de la philosophie humaniste qu'un électronicien britannique a conçu en 2005 le Mosquito, dispositif de harcèlement acoustique [*sic*] diffusant du 18 kHz à forte intensité dans le but assumé de faire fuir les groupes de jeunes antisociaux désœuvrés.

Pour terminer ce bref tour d'horizon des modalités sensorielles humaines, nous citerons simplement la gustation et l'olfaction qui, en captant les molécules aromatiques et odorantes, permettent la perception de l'information chimique.

2.3.3.2. Le système auditif humain

Nous allons à présent poursuivre notre *cycle de la parole*, en étudiant l'anatomie et la physiologie de l'audition au cours des différents étages des voies auditives, du pavillon de l'oreille externe à l'aire de Broca dans le cortex cérébral.

L'oreille

L'oreille humaine est un organe sensoriel complexe, capable d'une extrême précision et doté d'une mécanique sophistiquée. Nous pouvons la subdiviser en trois parties : externe (pavillon et conduit auditif), moyenne (tympan et osselets) et interne (cochlée et vestibule ; Figure 2.7.). Comme les oreilles gauche et droite ne présentent pas dans l'absolu de différences anatomo-fonctionnelles, nous parlerons de l'oreille et de ses différents éléments au singulier dans les paragraphes suivants.

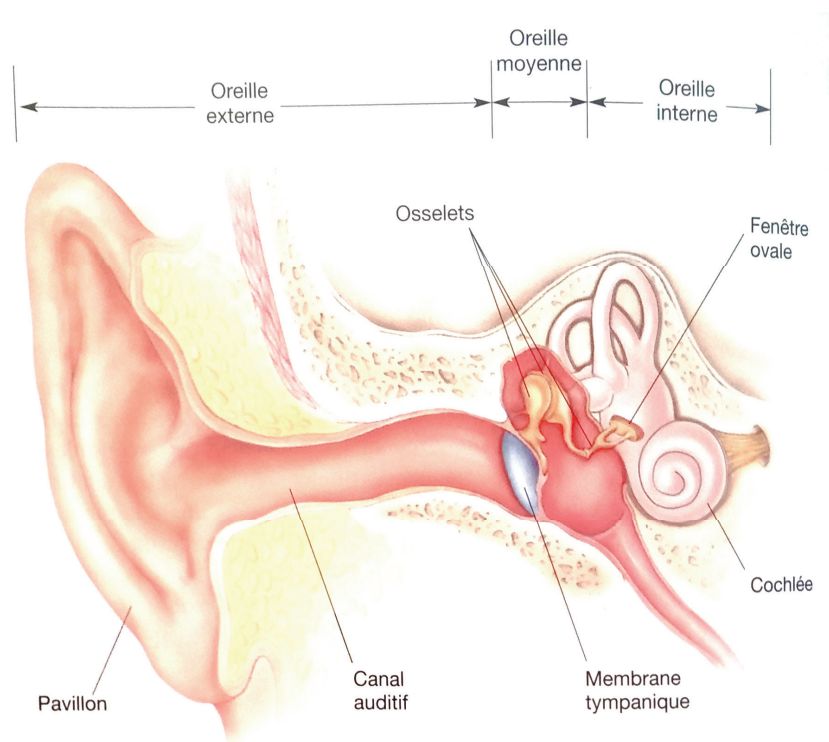


Figure 2.7. Anatomie de l'oreille dans sa globalité, avec ses trois parties externe, moyenne et interne (cette image est adaptée de l'excellent ouvrage *Neurosciences : à la découverte du cerveau*, 3^{ème} édition, Bear *et al.*, 2007).

Les ondes acoustiques véhiculant le message linguistique d'intérêt mêlé d'un bruit plus ou moins important sont d'abord filtrées par le torse et la tête, avant de parvenir à nos oreilles. Le pavillon, composante la plus visible de l'oreille externe, amplifie alors d'une dizaine de décibels (dB) les fréquences comprises entre 2 et 4 kHz, pertinentes pour la perception des phonèmes, grâce à sa forme globale en entonnoir. De plus, sa structure détaillée comporte de nombreuses circonvolutions, qui participent à la localisation de la source sonore en jouant un rôle de filtre linéaire sélectif à la direction : un même son en provenance d'une source supérieure ou inférieure aux oreilles sera réfracté par des reliefs différents au niveau du pavillon, ce qui lui confèrera une coloration spectrale différente. De cette manière, les pavillons impriment aux sons une empreinte d'oreille (*earprint*), et tout changement dans leur structure entraîne immédiatement une dégradation des performances de localisation sur l'axe vertical (*élévation* ; Figure 2.8.), sans influence sur la localisation dans l'axe horizontal (*azimuth*). Cependant, ces nouvelles empreintes, ou matrices d'associations entre spectres des sons et localisations, sont vite réappries, et les performances de localisation reviennent à leur niveau normal après un peu plus d'un mois (Hofman *et al.*, 1998).

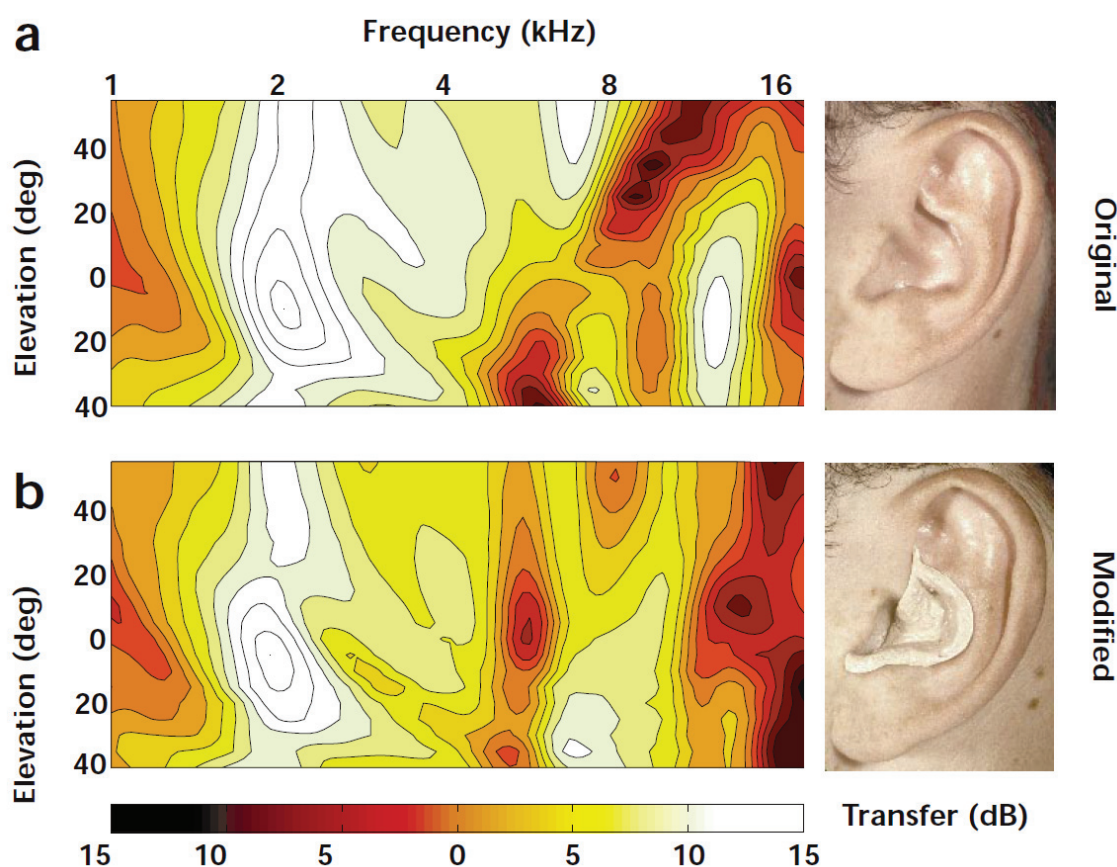


Figure 2.8. Fonctions de transfert pour un pavillon normal (en haut) ou modifié avec un moulage en silicone (en bas), obtenues en enregistrant le signal acoustique à l'aide d'un microphone placé au niveau du tympan, pour différentes valeurs d'élévation sur la ligne médiane (Hofman *et al.*, 1998). Le moulage change ainsi la fonction de transfert du pavillon, et donc le contenu spectral des sons perçus.

Une fois filtrées et amplifiées, les ondes sonores pénètrent dans le conduit auditif externe. Cette cavité tubulaire d'environ 2,5 cm de long, 8 mm de diamètre et en forme de S dans le plan horizontal, relie le pavillon au tympan. Le conduit auditif externe comprend une première section de nature cartilagineuse et fibreuse, et une seconde creusée dans l'os temporal. Il est recouvert d'une couche de peau décroissant en épaisseur (de 1 mm à proximité du pavillon à 0,2 mm au tympan), qui protège l'oreille moyenne de la poussière et autres germes indésirables grâce à la présence de poils et de nombreuses glandes sécrétant le cérumen. Aussi, cette peau se replie en cul-de-sac au niveau de la membrane tympanique, fait notable puisque cette configuration tissulaire ne s'observe nulle part ailleurs dans le corps humain. Du fait de ses capacités de résonance liées à sa forme incurvée, le conduit auditif externe amplifie d'environ 15 dB les fréquences entre 2 et 5 kHz (Wiener & Ross, 1946 ; Shaw, 1974).

Loin de n'avoir qu'une fonction esthétique ou punitive (en offrant une prise de choix sur les enfants irrévérencieux), l'oreille externe imprègne les ondes sonores d'une information sur l'élévation de leur source, et les amplifie d'environ 20 dB à des fréquences centrées sur 3 kHz et porteuses d'indices phonémiques pertinents pour l'intelligibilité de la parole.

Oreille moyenne

Le tympan est à l'interface entre le conduit auditif externe de l'oreille externe, et le marteau de l'oreille moyenne. Souple du fait des nombreuses fibres de collagène la composant, cette membrane conique d'une dizaine de millimètres de diamètre est comparée à une peau de tambour, ce qui explique son nom anglophone *eardrum*. Les ondes sonores transmises *via* le conduit auditif externe font entrer en vibration le tympan qui, raccordé au manche (*apophyse*) du marteau, retransmet les ondes sonores à la chaîne des osselets. Cette chaîne ossiculaire, constituée des trois plus petits os du corps qui sont le marteau (*malleus* ; 8 mm de long, pour une masse de 25 milligrammes), l'enclume (*incus* ; 7 mm, 27 mg) et l'étrier (*stapes* ; 3 mm, 4 mg), répercute et amplifie les vibrations tympaniques sur la fenêtre ovale cochléaire (Figure 2.9., de gauche à droite). L'amplification d'environ 60 dB est la conséquence de l'adaptation d'impédances entre milieux aérien canalaire externe et liquidien cochléaire interne¹, par le double jeu d'un effet levier entre marteau et enclume, et de la concentration de l'énergie en un minuscule point d'impact de l'étrier sur la fenêtre ovale (d'une surface de 3 mm², par rapport aux 60 mm² du tympan).

Les trois osselets sont maintenus dans la cavité remplie d'air nommée *caisse du tympan* par un système de ligaments et de deux muscles : le *tenseur du tympan* et le *stapédien* (plus petit muscle du corps humain, avec ses 2 mm ; Figure 2.9.). Ce dernier intervient dans le

¹ Les ondes sonores se propagent plus facilement dans le milieu aérien que dans le milieu liquidien. L'impédance du milieu liquidien est donc plus grande, ce qui signifie que ce milieu oppose une plus grande résistance au passage des ondes. Sans adaptation d'impédance, soit sans l'amplification des vibrations tympaniques, la plupart des ondes transmises à la cochlée seraient trop faibles pour être détectées par les cellules ciliées sensorielles.

réflexe acoustique ou *réflexe stapédien*, pour protéger l'oreille interne des sons de trop forte intensité (au-delà de 80 dB SPL²), à même de causer des dommages irréversibles au niveau des cellules sensorielles de la cochlée. Pour ce faire, le muscle stapédien éloigne l'étrier de la fenêtre ovale marquant l'interface avec l'oreille interne, pour atténuer d'environ 15 dB l'énergie vibratoire transmise. Cependant, ce réflexe met au moins 10 voire 20 millisecondes à produire son effet, ce qui laisse l'oreille vulnérable aux sons impulsionnels de très forte intensité. Aussi, il est également activé lors de la production de parole, pour limiter l'effet de masquage des basses fréquences des sons produits sur l'intelligibilité de la parole externe concomitante (Borg, 1975). Notons enfin que chez l'animal non-humain, le tenseur du tympan intervient aussi dans cette boucle réflexe, en rigidifiant (en *tendant*) la membrane tympanique, réduisant ainsi son excitabilité par les ondes acoustiques.

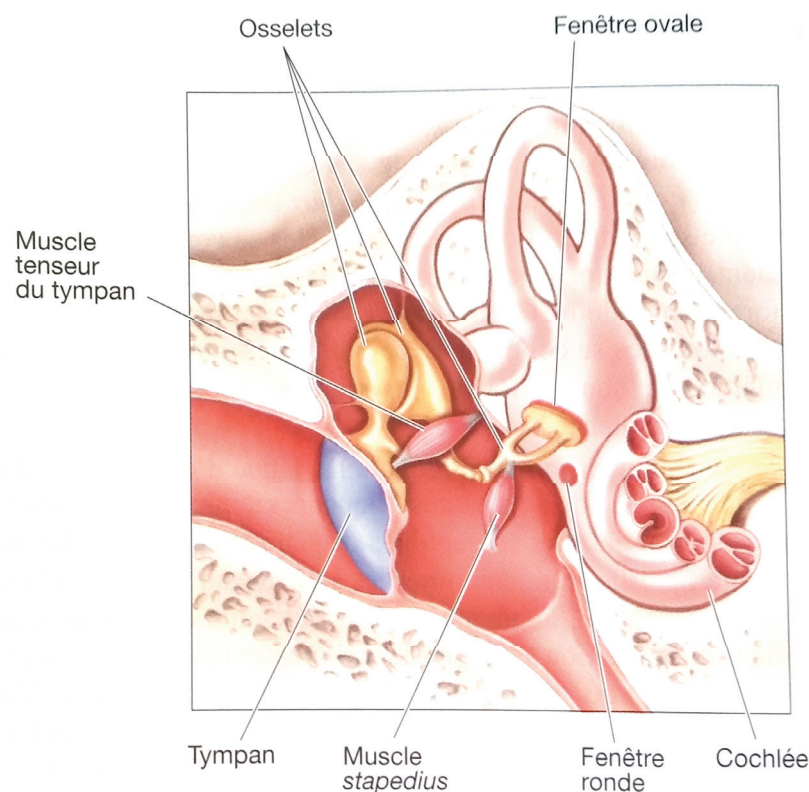


Figure 2.9. Schéma l'oreille moyenne avec ses principaux constituants (Bear *et al.*, 2007). Le son arrive par le canal auditif, sur le schéma à gauche du tympan. Les osselets, de gauche à droite, sont le marteau, l'enclume et l'étrier. A droite, une coupe de la cochlée est présentée (en comparaison à la Figure 2.7.). Enfin, la trompe d'Eustache est représentée en bas du schéma.

L'air remplissant la caisse du tympan est à une certaine pression qui, si différente de la pression atmosphérique externe, entrainera une déformation concave ou convexe de la

² SPL (*Sound Pressure Level*) : intensité physique, objective. S'oppose au dB SL (*Sensation Level*), qui mesure le volume sonore perçu, subjectif ; et aux dB HL et nHL (*Hearing Level* et *normal-HL*), valeurs normalisées sur une population et utilisées respectivement dans l'audiométrie tonale ou vocale et dans les PEAP (ou Click ABR ; voir chapitre 3).

membrane tympanique (en fonction de la direction du différentiel de pression), donc une réduction de ses capacités vibratoires et ultimement une baisse d'audibilité des ondes sonores. L'harmonisation des pressions se fait au travers de la trompe d'Eustache, reliée au pharynx et s'ouvrant lors des bâillements, de la déglutition ou de *manœuvres d'équilibrage* bien connues des plongeurs. Ce conduit offre aussi une fonction de drainage de l'oreille moyenne vers le pharynx.

Enfin, si le tympan, les osselets et la fenêtre ovale sont sollicités par les ondes acoustiques véhiculées par le conduit auditif externe, ils peuvent aussi retransmettre les vibrations osseuses de la boîte crânienne (*bone conduction*). C'est particulièrement le cas lorsque nous parlons, car les basses fréquences de notre voix sont mieux conduites par voie osseuse que par voie aérienne ; cela permet de comprendre notre étonnement lorsque nous écoutons notre voix enregistrée qui, sans ces basses fréquences seulement audibles par le locuteur, paraît alors bien creuse et aigüe.

Oreille interne - structure générale

Cavités osseuses creusées dans le rocher de l'os temporal, le vestibule et la cochlée sont les deux composantes de l'oreille interne. Le vestibule, impliqué dans le sens de l'équilibre, possède trois canaux semi-circulaires qui lui permettent d'approcher le fonctionnement d'un gyroscope. Il ne participe cependant pas à l'audition, aussi le laisserons-nous de côté.

La quintessence de l'ingénierie du corps humain, cerveau mis à part, se trouve peut-être dans la cochlée. Celle-ci doit son nom à sa forme spirale et conique, rappelant une coquille d'*escargot* : *cochlea* en latin, tiré du grec *κοχλίας* (lire « *kochlias* »). Aussi appelée limaçon, cet organe sensoriel haut de 4 mm s'enroule sur deux tours et demi de sa base à son sommet (*apex*), autour d'un axe osseux appelé *columelle*, et sa section se divise en trois compartiments principaux : la rampe vestibulaire, le canal cochléaire (ou rampe médiane, *scala media*) et la rampe tympanique (Figure 2.10., pour l'ensemble de cette description). Les rampes vestibulaire et tympanique sont toutes deux remplies d'un liquide, la *pérylymphe*, et ne communiquent entre elles qu'en une ouverture située au niveau de l'*apex* : l'*hélicotrème* (composé des mots grecs *ἑλῑξ* [*hélix*] signifiant *spirale*, et *τρῆμα* [*trêma*] signifiant *perforation* ou *trou*). A l'exception de cette ouverture, elles sont parfaitement séparées par la lame spirale osseuse et le canal cochléaire en prolongement de celle-ci. Le canal cochléaire est quant à lui rempli d'un liquide, l'*endolymphe*, et est entouré, comme nous venons de le voir, par les rampes vestibulaire et tympanique. Il est séparé de la première par la membrane de Reissner, simple barrière ionique séparant endolymphe et pérylymphe, et de la seconde par la membrane basilaire, sur laquelle repose l'organe de Corti.

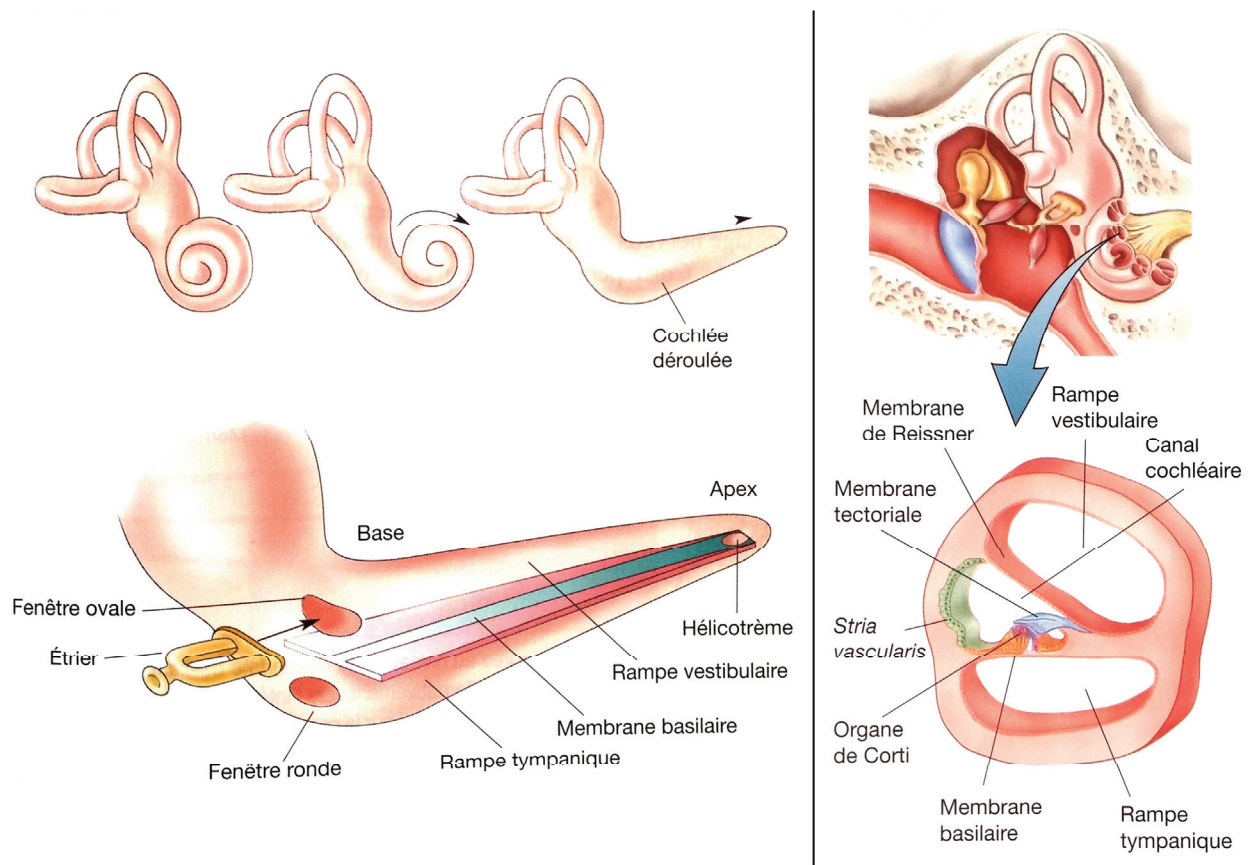


Figure 2.10. Représentation de la cochlée déroulée (à gauche) et de ses différentes composantes, dont l'organe de Corti (à droite ; Bear *et al.*, 2007).

A la base de la cochlée, la rampe vestibulaire (aussi appelée rampe montante) est séparée de l'oreille moyenne par la fenêtre ovale, et la rampe tympanique (descendante) par la fenêtre ronde. Les vibrations de l'étrier sur la fenêtre ovale entraînent une déformation de celle-ci vers l'intérieur de la cochlée, déformation qui s'observe immédiatement dans le sens opposé au niveau de la fenêtre ronde, du fait de l'incompressibilité de la périlymphe, et de la communication entre rampes montante et descendante au travers de l'hélicotreme. L'onde acoustique se propage dans la périlymphe, et fait entrer en vibration la membrane basilaire et son organe de Corti en un endroit précis de la spirale, fonction de la fréquence de l'onde.

Une image classiquement donnée pour schématiser la membrane basilaire est celle d'un clavier de piano enroulé sur lui-même. Pour se rapprocher de la réalité, nous pouvons imaginer ce clavier à l'envers, c'est-à-dire commençant par les aigus et terminant par les graves. Et pour pouvoir mieux comprendre cette observation, nous pouvons imaginer ce clavier en forme de palme : étroite (150 μm) et rigide à la base, et large (450 μm) et souple au sommet (Figure 2.11.). Soit dit en passant, et sauf le paramètre de rigidité, la membrane basilaire serait mieux représentée par un clavier de xylophone, vibraphone, marimba ou glockenspiel, que de piano (dont toutes les touches ont la même dimension, comme ce sont les cordes sous-jacentes qui diffèrent en diamètre, longueur et tension). Cette représentation des fréquences au long de la

membrane basilaire (des hautes fréquences à la base jusqu'aux basses à l'apex) correspond à la *tonotopie*, que nous allons par la suite retrouver à presque tous les étages du système nerveux auditif humain.

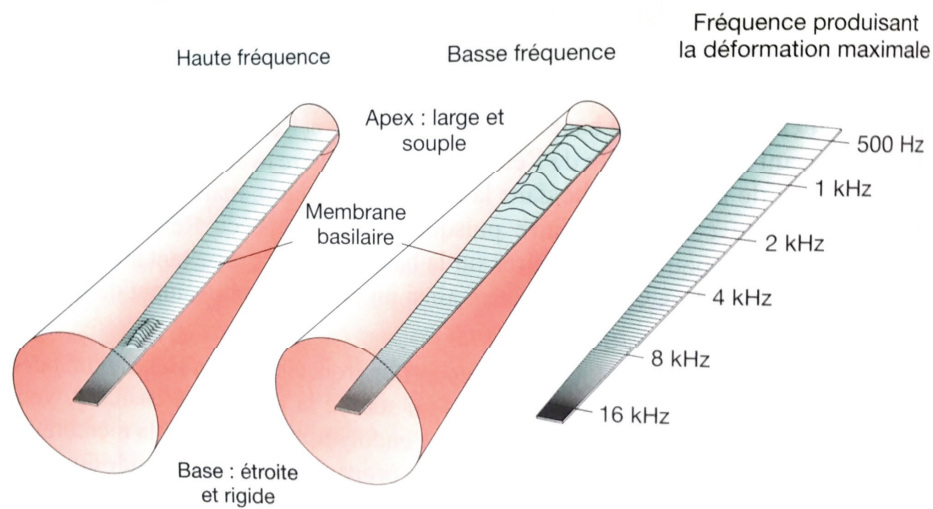


Figure 2.11. Membrane basilaire et origine de la tonotopie (Bear *et al.*, 2007). La fenêtre ovale, « l'entrée » du signal sonore dans la cochlée, se situe au niveau de la base. L'hélicotreme n'est pas représenté ici.

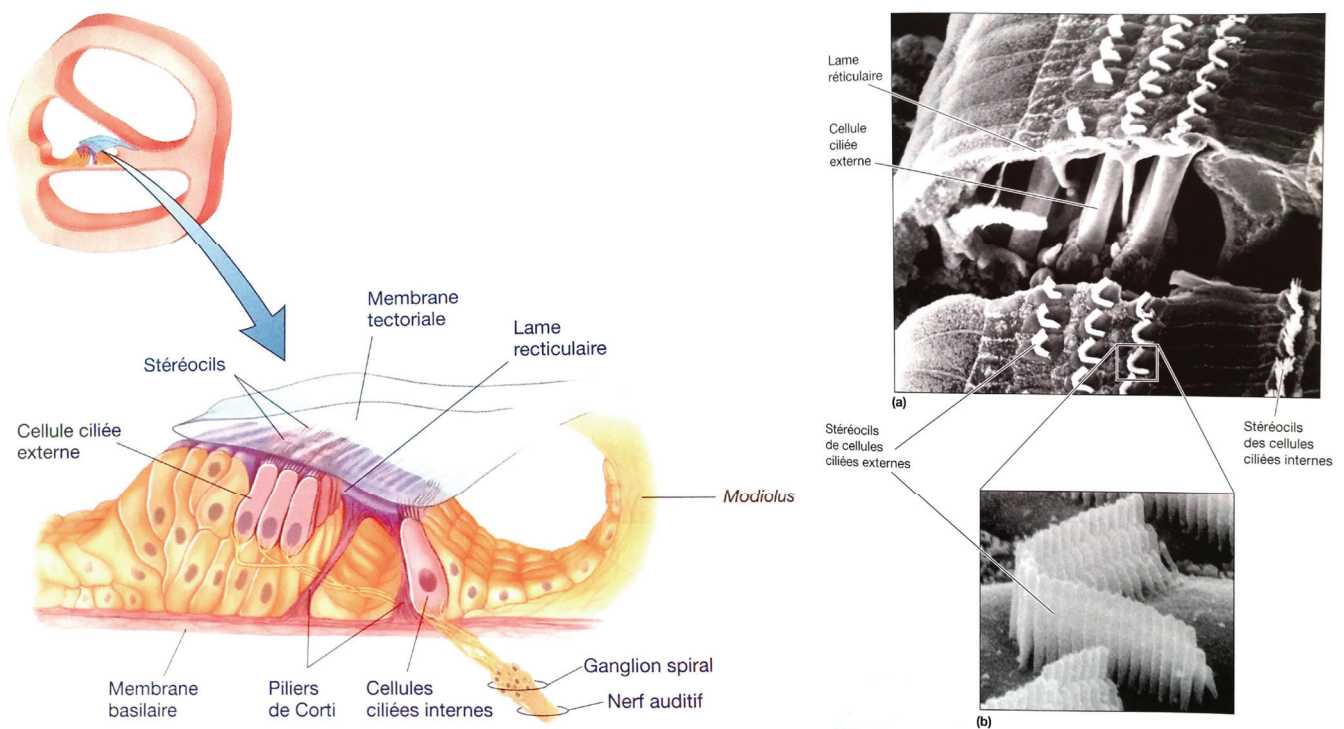


Figure 2.12. Schéma de l'organe de Corti et vue en microscopie électronique à balayage (à droite ; Bear *et al.*, 2007).

L'organe de Corti peut être considéré comme le cœur de l'oreille-organe sensoriel, en cela qu'il assure la transition entre système sensoriel et système nerveux, entre cellules ciliées

internes et neurones du nerf auditif, entre information mécanique et information électrique. Posé sur la membrane basilaire du côté canal cochléaire et proche de la lame spirale (Figure 2.12., en haut à gauche), l'organe de Corti consiste en une juxtaposition d'une rangée de 3500 cellules ciliées internes (CCI) et de trois à cinq rangées totalisant 12000 cellules ciliées externes (CCE), tout du long de la membrane basilaire (Møller, 2000). Les cellules ciliées sont appelées ainsi du fait de la présence d'une touffe de stéréocils (*hair bundle*) en leur sommet, qui a pour propriété essentielle sa mécano-sensibilité. Cette touffe comprend environ 60 stéréocils disposés en U pour les CCI, et entre 50 et 150 stéréocils disposés en V ou W pour les CCE (Figure 2.12., à droite). CCI et CCE sont entourées par des cellules de soutien, et séparées par le triangulaire *tunnel de Corti* formé par les piliers de Corti ; aussi, le sommet des cellules ciliées et des cellules de soutien forment la lame réticulaire, membrane résistante soutenue par les piliers de Corti et séparant périlymphe et endolymphe (Figure 2.12., en bas à gauche). Les touffes des CCI et CCE dépassent de cette lame pour baigner dans l'endolymphe, à l'intérieur du canal cochléaire, tandis que les corps cellulaires de ces mêmes CCI et CCE baignent dans la périlymphe. Enfin, une membrane gélatineuse, la membrane tectoriale, surplombe les CCI et intègre les cils des CCE (et seulement des CCE) en leur sommet. Remarquons la position des CCI et CCE à l'interface endolymphe/périlymphe qui, du fait de l'important gradient de concentration ionique entre ces deux liquides, conditionne les mouvements ioniques déclenchant la transduction du signal mécanique en signal électrique. Tapissant la face latérale du canal cochléaire, la stria vasculaire (*stria vascularis*) régule les concentrations ioniques des deux lymphes, pérennisant de fait le fonctionnement cochléaire.

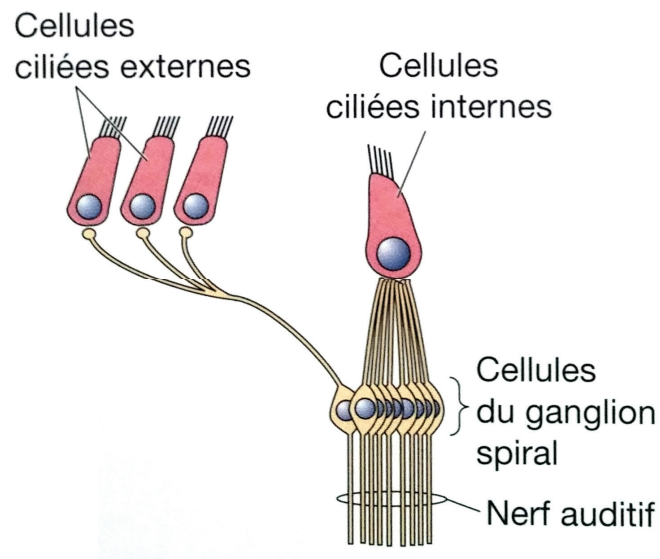


Figure 2.13. Innervation des cellules ciliées internes et externes (Bear *et al.*, 2007).

Les CCI sont les véritables cellules sensorielles de la cochlée, car ce sont elles qui transmettent l'information acoustique aux voies nerveuses auditives afférentes (conduisant

l'information de la périphérie au centre). Elles sont innervées à leur base par 95% des fibres du nerf auditif, chaque CCI étant connectée à une vingtaine de neurones afférents dont le corps cellulaire se trouve dans le ganglion spiral (Figure 2.13.). Les CCE reçoivent surtout en ce qui les concerne une innervation efférente, en étant ainsi sous contrôle du *système efférent olivo-cochléaire médian*. De plus, elles possèdent de nombreuses unités d'une protéine transmembranaire, la *prestine*, qui change de configuration et donc de taille lors de la dépolarisation cellulaire, leur conférant une capacité d'*électromotilité* (Mellado *et al.*, 2008). Leurs mouvements contractiles émettent d'ailleurs de très petits sons, enregistrables depuis le conduit auditif externe : les otoémissions acoustiques.

Oreille interne - physiologie

Maintenant que nous avons décrit la structure extrêmement sophistiquée de la cochlée, nous allons étudier son fonctionnement. Imaginons pour cela un son pur à une fréquence donnée, qui arrive à la fenêtre ovale. L'onde acoustique sinusoïdale va se propager dans la périlymphe et faire osciller la membrane basilaire à un endroit fonction de sa fréquence (proche de la base s'il s'agit d'un son haute-fréquence, et de l'apex si sa fréquence est basse).

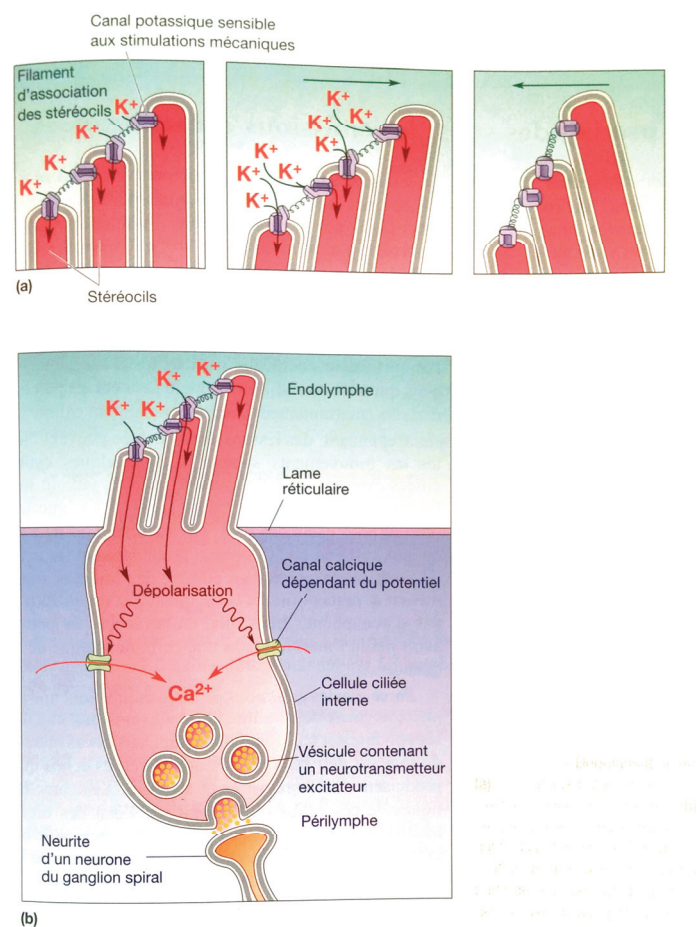


Figure 2.14. Voie de signalisation mécano-sensible des cellules ciliées (Bear *et al.*, 2007).

Si l'onde est suffisamment ample, c'est-à-dire si le son est suffisamment intense, alors le mouvement de la membrane basilaire fera entrer les cils des CCI en contact avec la membrane tectoriale : ceux-ci vont se plier et la contrainte mécanique ainsi imposée va entraîner l'ouverture de canaux ioniques mécano-sensibles. L'entrée massive d'ions potassium K^+ et calcium Ca^{2+} dans la CCI va entraîner sa dépolarisation, et la fusion de vésicules au niveau de sa partie basale pour externaliser des neurotransmetteurs, qui transmettent aux afférences neuronales le signal sonore (Figure 2.14.). La comparaison au piano reprend ici tout son droit, avec la membrane basilaire comme touches, les CCI comme marteaux, et les fibres du nerf auditif comme cordes.

Les CCE, directement liés à la membrane tectoriale, subissent la même contrainte mécanique menant à leur dépolarisation, mais cette dépolarisation a sur elles un tout autre effet : elle entraîne le changement de configuration de la prestine qui, en se contractant, réduit la taille de la CCE. La membrane tectoriale est alors rapprochée des cils des CCI, amplifiant ainsi leur réception du signal vibratoire, ce qui est indispensable pour la perception des sons peu intenses (Hudspeth, 2014). Aussi, cette amélioration de la sensibilité des CCI s'accompagne d'une augmentation de la sélectivité fréquentielle, due à la localisation précise de l'activité des CCE le long de la membrane basilaire. Tout le processus décrit se reproduit à chaque cycle de l'onde sonore, si bien que les neurones afférents recevront un signalement par cycle. Notons enfin que l'action des CCE peut également être médiée par le système efférent olivo-cochléaire médian : ceci peut sous-tendre des processus attentionnels où l'on se préparerait à entendre un son à une certaine fréquence, en faisant se contracter les CCE pour augmenter la sensibilité des CCI, autrement dit les faire réagir à un son de moindre intensité.

En résumé, la cochlée transforme le signal acoustique qu'elle reçoit par sa fenêtre ovale en signal électrique généré dans les neurones connectés aux CCI. Cette transformation se fait au niveau de l'organe de Corti, posé sur la membrane basilaire qui s'enroule tout du long de la cavité spirale osseuse qu'est la cochlée. Du fait de ses propriétés de largeur et de rigidité variant graduellement de la base à l'apex cochléaire, la membrane basilaire entre en vibration à des endroits précis pour des fréquences données : cette association entre fréquence et endroit s'appelle tonotopie. Au niveau de la vibration, les CCI entrent en collision avec la membrane tectoriale, ce qui entraîne une pliure de leurs stéréocils et l'ouverture de canaux ioniques sensibles à l'étirement ; un flux massif d'ions potassium K^+ entre alors dans la cellule ciliée, grâce au gradient ionique important entre endolymphes et périlymphe, et entretenue par la strie vasculaire. Dans la CCI, une chaîne de signalisation cellulaire entraîne *in fine* la libération du neurotransmetteur *glutamate* à la base, au niveau des synapses formées avec les neurones afférents du nerf auditif. Notons que ce flux ionique génère un champ électrique, enregistrable par des électrodes placées à proximité de la cochlée voire même au niveau du scalp, et dénommés potentiels microphoniques cochléaires (*cochlear microphonics*).

Voies nerveuses auditives sous-corticales

Recevant les afférences des CCI, les fibres du nerf auditif transmettent l'information auditive au noyau cochléaire ipsilatéral, qui projette un faisceau de fibres jusqu'au complexe olivaire supérieur controlatéral. L'information transite ensuite principalement du même côté, par les noyaux du lemnisque latéral jusqu'au colliculus inférieur, dans le mésencéphale (Figure 2.15.). Ce parcours est le schéma classiquement présenté dans le but de simplifier une réalité éminemment plus complexe, avec à l'œuvre de nombreux neurotransmetteurs, synapses, types cellulaires, subdivisions de noyaux, connexions croisées entre mêmes niveaux, niveaux successifs et même niveaux non-consécutifs, connexions pouvant de plus être excitatrices ou inhibitrices, et ascendantes ou descendantes ! Toute cette complexité est à garder à l'esprit pour nuancer la portée des affirmations suivantes.

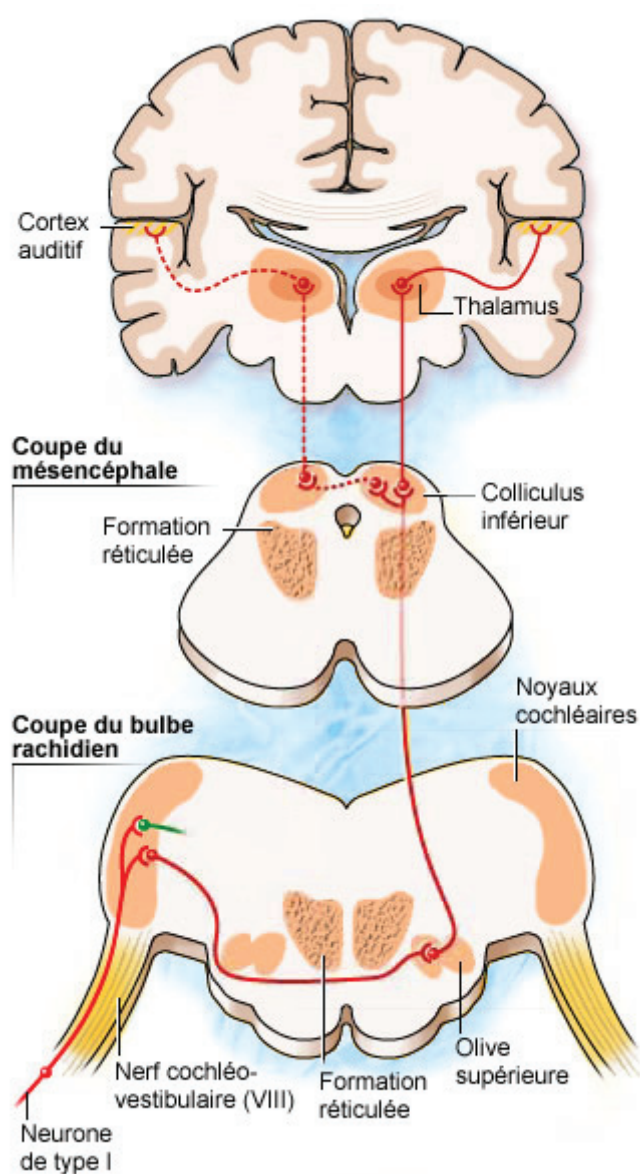


Figure 2.15. Voies auditives primaires ascendantes : schéma simplifié. La lecture se fait de bas en haut.

Le nerf auditif est composé de neurones bipolaires qui connectent les cellules ciliées internes avec leurs dendrites, et le noyau cochléaire avec leurs axones. Leurs corps cellulaires (soma) forment le ganglion spiral près de l'axe de la cochlée. Aussi, les fibres sont organisées en conservant la tonotopie de la membrane basilaire, avec les fibres connectées à l'apex et représentant les basses fréquences au centre, autour desquelles s'enroulent les fibres codant pour les moyennes et hautes fréquences.

Le noyau cochléaire est divisé en deux parties, ventrale et dorsale, la partie ventrale étant elle-même divisée en deux parties, antérieure et postérieure. L'intégralité des projections axonales des neurones constituant le nerf auditif font synapse dans le noyau cochléaire antéro-ventral, où des cellules globulaires de type « bushy » ont la capacité de transmettre le signal avec une grande fidélité temporelle. Des projections sont envoyées majoritairement de l'autre côté de la ligne médiane, et contactent le complexe olivaire

supérieur, le lemnisque latéral et le colliculus inférieur (voir ci-après). L'information arrivant au noyau cochléaire antéro-ventral transite également dans la division postéro-ventrale, où des neurones de type « pieuvre » (*octopus*) ont pour propriété de synchroniser leurs décharges avec la phase du *stimulus* acoustique ; leurs fibres myélinisées contactent le lemnisque latéral par une synapse de type *calyce de Held* offrant une libération massive de neurotransmetteurs sur une large zone de contact avec le neurone post-synaptique, et permettant de ce fait une transmission extrêmement rapide et précise des informations auditives. Aussi, des neurones « chopper » encodent le début du signal acoustique, sans rapport particulier à sa structure temporelle.

Le complexe olivaire supérieur est une constellation de neuf noyaux neuronaux, les trois principaux étant les noyaux latéral, médian, et médian du corps trapézoïde. Les noyaux médian et latéral regroupent les flux d'information provenant des deux noyaux cochléaires, pour opérer une détection de coïncidence sous-tendant la localisation spatiale de l'azimut (plan horizontal). En effet, l'onde acoustique est différemment filtrée par le buste et la tête en fonction de la position de la source sonore par rapport à l'auditeur. L'oreille ipsilatérale à la source recevra donc plus tôt une information acoustique plus ample, car moins filtrée notamment au niveau des hautes fréquences. Le noyau médian du complexe olivaire supérieur compare alors le temps d'arrivée de l'information auditive, et les différences d'intensité sont comparées par le noyau latéral. Ce mécanisme ne fonctionne logiquement pas pour des sources localisées sur la ligne médiane et donc équidistantes des deux oreilles ; la localisation d'une source sur la ligne médiane est en partie résolue par le filtrage acoustique du pavillon de l'oreille externe, que nous avons vu précédemment.

Le lemniscus latéral possède trois parties, dorsale, ventrale et intermédiaire, projetant sur le colliculus inférieur de façon principalement inhibitrice avec le recours massif au neurotransmetteur GABA. Il est reçoit l'information acoustique des deux oreilles pour faire la jonction entre tronc cérébral et mésencéphale, et est impliqué dans son traitement temporel et de modulation en amplitude.

Le colliculus inférieur, divisé en quatre parties (division centrale, cortex dorsal, noyau dorso-médian et noyau latéral), est fréquemment décrit comme le centre de tri, le hub par lequel transitent presque tous les flux ascendants et descendants d'information auditive. Il intègre donc les informations des différents noyaux sous-corticaux dont il reçoit les projections, et projette lui-même sur le thalamus auditif pour les voies ascendantes, et sur tous les niveaux inférieurs jusqu'au noyau cochléaire. Il intègre aussi des informations multi-sensorielles, notamment visuelles de par sa proximité avec le colliculus supérieur. Sa division centrale possède une structure en couches, chacune dédiée au traitement d'une bande de fréquences, et est impliquée dans l'intégration de l'information binaurale pour la perception de la hauteur de son (pitch), ainsi que de la localisation spatiale en regroupant les informations horizontales et verticales (azimut et élévation).

Retenons de cette ascension quelques notions principales dont nous aurons usage par la suite : à son arrivée dans le noyau cochléaire antéro-ventral, le signal peut être efficacement envoyé au colliculus inférieur par des cellules « bushy », ou passer à la division postéro-ventrale pour être envoyé par des neurones « octopus », ces deux types cellulaires permettant la transmission d'information temporelle fine, précisément liée au *stimulus*. Ajoutons que la limite de verrouillage de phase (*phase-locking*) décroît au fur et à mesure de la montée dans les voies auditives nerveuses, partant d'une valeur d'environ 5000 Hz au niveau du nerf auditif, descendant entre 1500 et 3000 Hz dans le noyau cochléaire, pour tomber autour de 1000 Hz dans le colliculus inférieur. Aussi, l'information tonotopique liée à la structure cochléaire est conservée au niveau du colliculus inférieur.

Thalamus et cortex auditif

Avant d'arriver au cortex auditif primaire, l'information auditive en provenance du colliculus inférieur passe obligatoirement par le corps genouillé médian (subdivisé en parties ventrale, médiane et dorsale) du thalamus. La partie ventrale intègre de façon complexe les dimensions de fréquence, de latence et d'intensité du son. La plupart des neurones la constituant sont sensibles aux informations binaurales, et varient dans leur fonctionnement en fonction de l'état attentionnel.

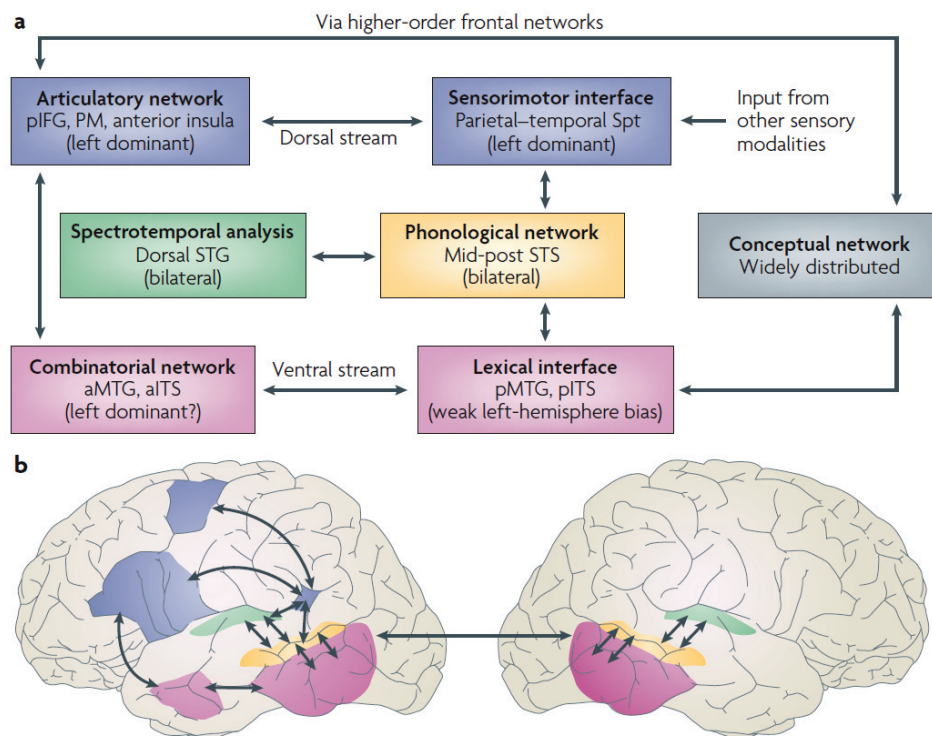


Figure 2.16. Traitement de l'information auditive dans le cortex cérébral (modèle à 2 voies ; Hickok & Poeppel, 2007). Le cortex auditif primaire est représenté en vert, la très célèbre aire de Wernicke en jaune orangé, et la non moins fameuse aire de Broca en indigo (sur la vue sagittale en bas à gauche, il s'agit de la zone la plus ample, dans le gyrus frontal inférieur postérieur, ou pIFG tel que dans le cadre en haut à gauche).

Le cortex auditif primaire se situe au niveau du gyrus de Heschl, dans la partie supérieure du lobe temporal. Son organisation en colonnes corticales garde l'information de tonotopie, et offre une représentation précise des sons en combinant fréquence, durée, intensité et timbre. Sa limite de suivi de la phase est communément fixée à 100 Hz environ (Steinschneider *et al.*, 2013).

L'information auditive continue ensuite son cheminement dans les aires auditives corticales secondaires et associatives, sa propagation s'arrêtant plus ou moins tôt en fonction de sa complexité, en termes de richesse informationnelle et de pertinence comportementale. Les sons de parole évoquent justement les réponses les plus complexes, avec une extraction de l'information phonémique maintenue en mémoire et combinée pour former les syllabes, les mots et les phrases.

Nous pouvons développer ici la notion de non-linéarité à l'ouvrage dans le traitement cortical auditif de la parole, en empiétant légèrement sur la partie suivante : une syllabe isolée, par exemple, ne produira pas la même réponse neuronale que si elle s'accompagne d'une autre syllabe pour former un mot. Intuitivement, nous pouvons nous représenter toutes les pensées qu'occasionne l'écoute d'un mot ; elles ne seraient évidemment pas présentes à l'écoute séparée de chacune des syllabes composant ce mot. « Le tout est plus que la somme des parties », célèbre maxime d'Aristote, trouve ici une illustration bien concrète. Dans la même logique que la détection de coïncidences de l'olive supérieure, différentes informations doivent arriver dans un certain ordre pour déclencher une réponse particulière. Pour la parole, la représentation spectro-temporelle au niveau du cortex auditif primaire va être « lue » par le réseau phonologique dans le *sulcus temporal supérieur* postéro-médian, qui en extraira une information abstraite : le phonème (Liebenthal *et al.*, 2005). L'activité se propagera ensuite le long de deux flux, ventral pour l'accès au message linguistique avec la sollicitation de l'aire de Wernicke, et dorsal pour l'accès aux informations articulatoires avec notamment l'aire de Broca (Hickok & Poeppel, 2007). Là encore, un phonème isolé n'activera pas les mêmes aires que plusieurs phonèmes associés, qui à leur tour n'activeront pas les mêmes aires que si cette association forme un mot doté de sens. De tels processus non-linéaires sont omniprésents, et doivent être gardés à l'esprit en étudiant les voies auditives humaines, en particulier dans le cas de l'encodage de la parole.

2.3.4. Cognition

Après avoir exposé les différentes structures et fonctions permettant d'acheminer l'information auditive au cerveau humain, nous allons à présent nous intéresser au traitement

de cette information. Bien entendu, cognition et perception ne sont pas séparées d'une frontière nette et infranchissable : il s'agit plutôt d'un *continuum* entre matière et esprit, entre physique et phénoménologie.

Le processus que nous allons décrire maintenant est un élément clé pour notre capacité à comprendre la parole. En effet, si l'information reçue par le cortex auditif a déjà subi plusieurs traitements aux niveaux hiérarchiquement inférieurs, elle n'en reste pas moins d'ordre acoustique. Or, l'acoustique de la parole est hautement variable, entre les locuteurs du fait de l'anatomie idiosyncratique de l'appareil phonatoire, pour un même locuteur du fait de l'impossibilité de répéter à l'identique la même séquence motrice, et dans les deux cas du fait du bruit constamment variable qui se mêle à la parole pendant sa transmission. Nous comprenons pourtant le message linguistique véhiculé par la parole, et ce quelles que soient les circonstances (tout du moins jusque dans une certaine mesure de clarté du locuteur et d'audibilité du signal par rapport au bruit) : ce constat suggère un processus cérébral capable de s'affranchir de cette variabilité acoustique non pertinente pour le message linguistique, afin d'extraire du flux sonore, variable et continu, une information linguistique, stable et discrétisée.

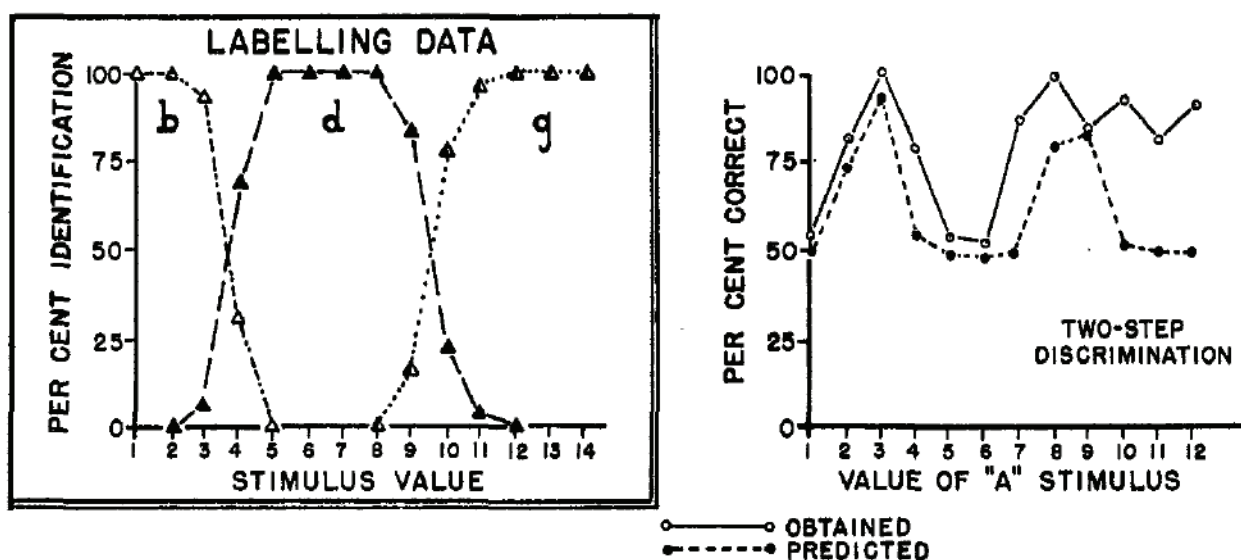


Figure 2.17. Découverte de la perception catégorielle (adapté de Liberman *et al.*, 1957). A gauche, courbe d'identification phonémique des 14 *stimuli* synthétiques du *continuum* /be/-/de/-/ge/ pour un participant : les frontières /be/-/de/ et /de/-/ge/ se situent respectivement entre les *stimuli* 3 et 4, et 9 et 10. A droite, courbe de discrimination phonémique (en traits pleins) : le participant identifie très clairement les *stimuli* 2 et 4 comme différents, alors qu'il est au niveau de la chance pour les *stimuli* 5 et 6 appartenant à la catégorie phonémique /de/.

En 1957, Alvin Liberman demande à des participants d'indiquer si des paires de syllabes sont composées de *stimuli* identiques, ou différents. Pour cela, il utilise un *continuum* de syllabes, c'est-à-dire un dégradé acoustique régulier reliant plusieurs syllabes qui s'opposent

sur un même trait phonétique (en l'occurrence le lieu d'articulation avec un *continuum* /be/-/de/-/ge/). En présentant des paires séparées par un nombre de pas constant dans le *continuum*, il constate que les participants jugent davantage ces paires comme différentes lorsqu'elles se situent à proximité de la frontière phonémique (l'endroit dans le *continuum* où la perception bascule d'un phonème à l'autre). Le corollaire est que nous percevons moins les différences acoustiques lorsqu'elles opèrent à l'intérieur d'une catégorie phonémique : le concept de perception catégorielle voit ainsi le jour (Liberman *et al.*, 1957 ; Figure 2.17.).

Par ce phénomène, des répétitions acoustiquement variables d'un même phonème sont alors groupées, entrent dans une même catégorie, stable et discrétisée.

Ceci n'est cependant qu'un premier pas vers la compréhension du message linguistique : les phonèmes successifs peuvent être groupés en mots, dont seul l'accès au lexique mental (sorte de dictionnaire neuronal des mots appris ; Figure 2.16., voie ventrale) garantira le passage du son au sens (*sound-to-meaning*). Les concepts de mots pourront alors s'associer en suivant une certaine syntaxe, la grammaire, pour conjuguer leur signification et ultimement faire émerger à la conscience le message du locuteur.

Enfin, il existe de nombreux effets cognitifs liés à la parole. Nous allons très brièvement en décrire trois :

- l'effet McGurk consiste en une modification du percept lié à un *stimulus* acoustique par un *stimulus* visuel concomitant et contradictoire ; en des termes plus concrets, l'écoute d'une syllabe /ba/ simultanée à la vision d'une personne prononçant /ga/ pourra donner un percept /da/, par exemple ;
- l'effet de transformation verbale est une modification du percept suite à l'écoute répétée d'un même *stimulus* ; par exemple, en écoutant une même syllabe /ba/ répétée plusieurs fois par seconde pendant plusieurs minutes, certains individus entendront d'autres syllabes, voire des mots ou des sons non-langagiers ;
- l'effet de satiété verbale est aussi causé par la répétition d'un mot, cette fois-ci produit par l'auditeur même ; à force de prononcer le mot donné, la perception de ce mot change, il peut se vider de son sens et sonner comme un mot nouveau, donnant l'impression que l'on ne l'avait jamais considéré en détail ;

Si le dernier effet n'a pas de raison particulière d'être rencontré dans les études constituant la partie expérimentale du présent manuscrit, le deuxième a été très souvent rencontré, avec des participants entendant des mots aussi variés que « maman », « bateau », « banane » ou « cabane » en réponse à une stimulation répétée de la même syllabe /ba/ ; l'effet McGurk est possiblement rencontré lorsque les participants regardent un film pendant

l'enregistrement de l'activité cérébrale en réponse aux sons de parole, bien que les occasions pour qu'il se manifeste sont moins nombreuses par rapport au deuxième effet. Or, l'impact de ces effets sur les réponses électrophysiologiques n'est pas bien connu, comme nous allons le voir par la suite.

2.3.5. Boucler la boucle

Nous avons suivi la parole dans son cycle, de sa production dans l'organe phonatoire humain, à sa cognition dans les méandres du cortex cérébral. Pour refermer le cycle, la conscience du message du locuteur doit engendrer la décision de produire à son tour de la parole pour lui répondre, qui elle-même engendrera une activité dans les aires cérébrales motrices pour activer l'appareil phonatoire, ce qui nous ramène à notre point de départ : la boucle est bouclée.

2.4. La parole et ses imperfections

Si le cycle de la parole que nous venons de détailler est un phénomène central dans nos vies quotidiennes comme à une échelle plus large, il n'est pas infallible, et est bien souvent dérangé dans son bon déroulement. Nous ne nous en rendons pas toujours voire même pas souvent compte, du fait de l'action croisée de mécanismes attentionnels et de compensation (suppléance mentale), entre autres.

Au niveau de sa production, la prononciation de parole nécessite un contrôle moteur très fin pour que le son résultant soit compréhensible par l'auditeur, si bien qu'elle devient rapidement altérée par la fatigue, le stress ou la consommation d'alcool. Cela est toutefois compensé par la perception catégorielle, qui offre une marge de tolérance à la variabilité des sons prononcés, y compris anormalement. Aussi, des petites variations dans la mécanique de l'appareil phonatoire peuvent changer le contenu spectral : une déshydratation suite par exemple à un discours trop long, ou une blessure au niveau de la pointe de la langue, entraîneront une déformation des sons prononcés, par exemple sous forme d'un zozotement.

Au niveau de sa transmission, l'environnement dans lequel est prononcée la parole est très rarement baigné de silence. Aussi, il arrive fréquemment que les indices phonétiques soient masqués par un bruit inopportun.

Au niveau de sa perception, l'impressionnante mécanique de l'oreille et toute la chaîne de traitement qui s'en suit sont aussi relativement fragiles, comme nous allons le voir dans la partie suivante. Les cellules ciliées ne se renouvellent en effet pas une fois détériorées.

Au niveau de sa cognition, le percept de parole représente parfois davantage l'esprit de l'auditeur que la parole du locuteur : « chacun entend ce qu'il veut bien entendre ». Cela peut être vu comme un effet secondaire des mécanismes de compensation, à l'origine censés rendre plus robuste la communication.

Nous disons parfois que les mots nous manquent. En effet, nos idées ont parfois besoin du langage pour se dessiner plus nettement. Le lien entre pensées et parole n'est pas pour autant évident, et recouvre une imperfection majeure du cycle de la parole : pour communiquer, le locuteur concrétise sa pensée par des mots ; le message part aux oreilles de l'auditeur, qui le reçoit plus ou moins bien et, s'il parvient à l'extraire correctement, interprète le message avec son propre système lexical, ses propres concepts et définitions. Il y a très peu de chances en réalité que la pensée d'origine du locuteur et celle finale de l'auditeur soient identiques. En d'autres termes, une grande place est laissée à l'interprétation, d'autant plus quand les interlocuteurs ne se connaissent pas, voire ne partagent pas la même culture.

2.4. Pathologies liées et traitements associés

Comme dans tout système, la complexité augmente le nombre de failles potentielles. Le cycle de la parole n'y échappe pas, bien au contraire : des troubles de différentes natures et ampleurs peuvent altérer ses étapes de production (aphasies, bégaiement, paralysies faciales, pour en citer quelques-uns), de perception et de cognition. Nous ne nous concentrerons ici que sur les pathologies liées à la perception auditive.

Les multiples dysfonctionnements possibles de la capacité à percevoir l'information acoustique sont regroupés sous l'étiquette de *troubles de l'audition*. Ceux-ci peuvent affecter chacun des organes du système auditif humain, aussi peuvent-ils être de transmission lorsqu'ils touchent l'oreille externe ou moyenne, de perception lorsque la cochlée ou le nerf auditif sont atteints, ou mixtes lorsque les deux niveaux précités se combinent. Aussi, les surdités centrales recouvrent les pathologies liées aux voies nerveuses centrales, à partir du noyau cochléaire.

2.5.1. Surdités de transmission

Une malformation congénitale, l'aplasie, entraîne une absence de conduit auditif externe. Si la chaîne des osselets de l'oreille moyenne et la cochlée sont en bon état, ils sont difficilement mis en vibration. Les prothèses à vibration osseuse permettent généralement bien de transmettre l'information auditive à la cochlée, en palliant à l'absence de transmission aérienne normale.

Le tympan peut être soumis à de fortes variations de pression, entraînant parfois sa rupture. Les fréquences sont alors moins bien transmises aux osselets, et entraînent une gêne notable. Il est cependant doté de capacités de cicatrisation, qui évitent bien souvent le recours à une autre thérapeutique.

Dans le cas des otites séreuses, au niveau de l'oreille moyenne, l'équilibre des pressions entre la caisse du tympan et l'environnement extérieur est perturbé par la présence abondante d'un liquide, la glu, rendant ainsi plus difficile la mise en mouvement des osselets, et par conséquent l'audition. Une paracentèse (incision du tympan) ou la pose d'un aérateur transtympanique (le « yoyo »), permet alors l'évacuation de la glu.

2.5.2. Surdités de perception - Zoom sur la presbyacousie

Elles peuvent être cochléaires (ou neurosensorielles), lorsque seule la cochlée est atteinte, ou rétro-cochléaires, quand le nerf auditif présente des déficiences.

Du fait de sa structure extrêmement fine et complexe, la cochlée est sans doute l'organe du système auditif le plus touché par diverses pathologies. Pour la surdité profonde, aussi appelée *cophose*, des électrodes peuvent être placées dans la cochlée pour prendre la relève des cellules ciliées défaillantes, remplaçant le signalement neurochimique du signal sonore par une stimulation électrique directe. Cette technologie est l'implant cochléaire et permet de restaurer une fonction auditive suffisante pour comprendre et utiliser la parole, malgré une résolution fréquentielle bien moindre que celle d'une cochlée pleinement fonctionnelle.

D'une façon moins totale, le vieillissement peut progressivement en altérer la structure et la fonction de la cochlée. Notons tout de même que le vieillissement n'est pas considéré comme une pathologie, tout du moins pas encore puisque des chercheurs américains ont tout récemment tenté de l'avancer en tant que telle, auprès de l'administration en charge donner aux médicaments leur autorisation de mise sur le marché (Check Hayden, 2015).

La surdité liée au vieillissement physiologique de l'oreille est appelée *presbyacousie*. Potentialisée par les traumatismes sonores que subissent les oreilles au cours de la vie, elle se

caractérise par une perte des cellules ciliées externes préférentiellement sur la partie basale de la membrane basilaire, entraînant une baisse de l'activation des cellules ciliées internes et une moindre sélectivité fréquentielle. La strie vasculaire a également été pointée par certaines études, dans son déficit progressif à réguler efficacement le gradient ionique entre périlymphe et endolymphe grâce à ses pompes Na/K-ATPase, réduisant de fait le flux ionique entrant dans les CCI, et *in fine* l'excitabilité cochléaire ; ce mécanisme pathogénique pourrait être médié par des dommages oxydatifs aux mitochondries des cellules de ce tissu (Spicer & Schulte, 2005). Toutefois, si la presbyacousie agit avant tout au niveau cochléaire, il n'en reste pas moins qu'elle affecte également la fonction rétro-cochléaire (Ouda *et al.*, 2015).

Au niveau comportemental, le malentendant presbyacousique « entend mais ne comprend pas ». Cela vient de la meilleure conservation des basses fréquences, suffisantes pour détecter un son de parole mais insuffisantes pour en percevoir les détails. Un son filtré et débarrassé de ses hautes fréquences produira d'ailleurs un effet semblable chez le normo-entendant : par exemple, nous confondons volontiers les /s/, /f/, /ch/ dans nos conversations téléphoniques, car les fréquences sont absentes au-delà de 3500 Hz.

L'incapacité du patient presbyacousique à percevoir les hautes fréquences dégrade sa perception des consonnes, et ses capacités de débruitage du signal sonore basées sur cette même bande de fréquence. Il peut ainsi être particulièrement gêné en milieu bruyant, tendant alors petit à petit à éviter les situations « à risque » (restaurants, soirées animées, fêtes de familles, *et cetera*). Selon son tempérament, il pourra s'isoler progressivement pour éviter d'avoir à faire répéter ses interlocuteurs et ainsi être confronté au handicap. Or une vie sociale active à un âge avancé ne peut que retarder l'apparition d'autres pathologies du vieillissement, telles que la démence de type maladie d'Alzheimer.

Bien souvent, l'utilisation d'aides auditives permet de restaurer une intelligibilité convenable, et en conséquence une vie active. Cependant, le choix des appareils auditifs, leur réglage, la prédiction de succès et le suivi des bénéfices potentiels ne sont de nos jours que très rarement basés sur des mesures objectives. Cela explique peut-être le cinquième de patients équipés préférant laisser leurs appareils dans le tiroir (Kochkin, 2010). Une mesure précise et fiable de la fonction auditive du patient presbyacousique permettrait alors certainement d'abaisser ce taux d'échec, en marquant par la même occasion l'entrée dans l'ère de l'audiométrie objective.

Enfin, un nombre croissant d'études met en évidence les bénéfices d'un entraînement auditif, c'est-à-dire d'exercices perceptifs, cognitifs et attentionnels permettant, au sens figuré, de remuscler les voies auditives pour les rendre plus performantes dans l'extraction et le traitement des phonèmes, malgré les atteintes au niveau cochléaire ; autrement dit, *pour faire plus avec moins*.

3. Electrophysiologie de l'audition et autres outils de mesure

3.1. Techniques à disposition du clinicien et du chercheur

3.1.1. Vue d'ensemble

Les différentes étapes du cycle de la parole peuvent être pour la plupart mesurées, en faisant appel à de nombreuses techniques et domaines d'étude. Par exemple, la production physiologique de la parole est étudiée par la linguistique et ses sous-disciplines phonétique et phonologie, les neurosciences et sciences cognitives, la psychologie, la médecine et disciplines associées comme l'orthophonie, la physique et les mathématiques, mais aussi l'informatique, l'ingénierie et la robotique, du fait des multiples applications industrielles possibles. Les frontières entre ces domaines sont parfois plus conceptuelles que concrètes, car ceux-ci peuvent plutôt être vus comme des compétences, des grilles de lecture permettant d'approcher préférentiellement la production de parole sous un certain angle. La transmission des sons de parole est quant à elle étudiée principalement par la physique et sa sous-discipline l'acoustique, la bioacoustique à mi-chemin entre biologie et acoustique, ainsi également que par l'ingénierie. Ces listes non-exhaustives donnent un bref aperçu de l'interdisciplinarité intrinsèquement liée aux multiples facettes du cycle de la parole. Dans cette première partie du présent chapitre, nous détaillerons les techniques médicales et neuroscientifiques à l'usage dans l'étude de l'audition humaine pour la perception de la parole.

Comme nous l'avons vu au chapitre précédent, la parole est un phénomène mêlant variabilité temporelle très fine (avec des phonèmes et des syllabes de quelques dizaines et quelques centaines de millisecondes respectivement) et invariances articulatoires et perceptives (une grande variété de sons produits ou perçus pouvant être assimilés à un même phonème bien qu'étant acoustiquement différents), notamment par le fait de la perception catégorielle. Si l'investigation de l'audition de la parole doit donc user de techniques à haute résolution temporelle pour accéder à la dynamique fonctionnelle précise des processus concernés, elle peut aussi bénéficier de techniques de moindre résolution temporelle mais de meilleure résolution spatiale, pour étudier l'anatomie ou les processus intégratifs, moins proches du signal acoustique brut.

Dans ce dernier cas, les techniques d'imagerie métabolique telles que l'*Imagerie par Résonance Magnétique fonctionnelle* (IRMf) et la *Tomographie par Emission de Positons* (TEP) mesurent l'afflux sanguin cérébral pour en déduire les zones actives, par exemple suite à l'écoute de sons de parole. Leur résolution temporelle de l'ordre de la seconde ne permet pas d'approcher la dynamique temporelle d'encodage de ces sons, tout du moins pas au niveau phonémique, mais ces techniques apportent des informations essentielles sur l'identité des aires corticales et sous-corticales recrutées dans la perception de la parole, et leur séquence d'activation (Hickok & Poeppel, 2007). L'électrophysiologie mesure plus directement l'activité cérébrale, par l'enregistrement des champs électriques et magnétiques générés lors de la communication entre neurones (plus précisément lors des courants post-synaptiques ; Garnero, 2001), avec une échelle propre à chaque technique, pouvant aller du neurone individuel à une population de neurones, voire à une aire cérébrale entière. L'*électroencéphalographie* (EEG) et la *magnétoencéphalographie* (MEG) permettent ainsi de se placer à l'échelle temporelle de l'encodage neuronal des variations fines de la parole, en pâtissant toutefois d'une résolution spatiale moins bonne que les techniques d'imagerie métabolique précitées. Des méthodes plus ou moins sophistiquées de reconstruction de sources permettent parfois d'améliorer cet inconvénient, en localisant la ou les régions cérébrales génératrices du signal enregistré au niveau des capteurs, posés à la surface de la tête. Aussi, des électrodes sont parfois placées à l'intérieur même du cerveau humain, dans le cas de l'exploration neurologique de l'épilepsie ou de mesures peropératoires ; cette dernière technique, connue sous le nom d'EEG intracrânien ou EEG stéréotaxique (SEEG), offre alors une résolution spatiale très fortement augmentée, à l'échelle de l'électrode, au détriment néanmoins de la couverture spatiale, du fait du nombre limité d'électrodes implantées.

3.1.2. Outils d'exploration clinique des voies auditives

Dans un volet clinique, l'oto-rhino-laryngologie et sa sous-spécialité l'audiologie ont à disposition une panoplie de techniques pour caractériser les pathologies de l'audition. Comme nous l'avons vu au chapitre précédent, de nombreuses causes peuvent expliquer une surdité, de façon non-exclusive ; il est donc nécessaire pour traiter cette surdité d'en trouver la ou les causes, en utilisant pour cela des techniques étudiant le fonctionnement ou la structure du système auditif humain.

3.1.2.1. Comportement

Sur le plan comportemental, la psychoacoustique est une discipline mettant en relation les percepts de l'audition avec les paramètres acoustiques des *stimuli* évoquant ces percepts. Une batterie de tests de ce type éclaire le clinicien sur le déficit auditif probable.

L'audiométrie tonale définit les seuils de perception, en intensité acoustique relative à une population normale, en fonction des fréquences, la plupart du temps pour les octaves entre 250 et 8000 Hz. Le profil audiométrique obtenu donne une idée de la pathologie sous-jacente en fonction des fréquences atteintes : une perte sur les basses fréquences avec des hautes fréquences bien conservées est compatible avec une hyperpression des liquides de l'oreille interne ; une perte en encoche, à une fréquence précise, fera penser à un traumatisme sonore ou à une ototoxicité ayant entraîné la disparition des cellules ciliées sur une zone particulièrement sensible de la membrane basilaire (généralement autour du 3 et 6 kHz) ; enfin, l'observation de basses fréquences bien conservées avec une perte sur les hautes fréquences à partir de 1 ou 2 kHz sera compatible avec une presbyacousie (Figure 3.1.). Le caractère potentiellement symétrique des audiométries tonales pour chacune des deux oreilles informe aussi sur la nature du trouble. Ceux-ci peuvent se combiner, et il n'est pas rare d'observer une audiométrie de type presbyacousie pour une oreille, et un mélange de presbyacousie et de traumatisme sonore pour l'autre oreille, rendant alors asymétriques les fonctions cochléaires.

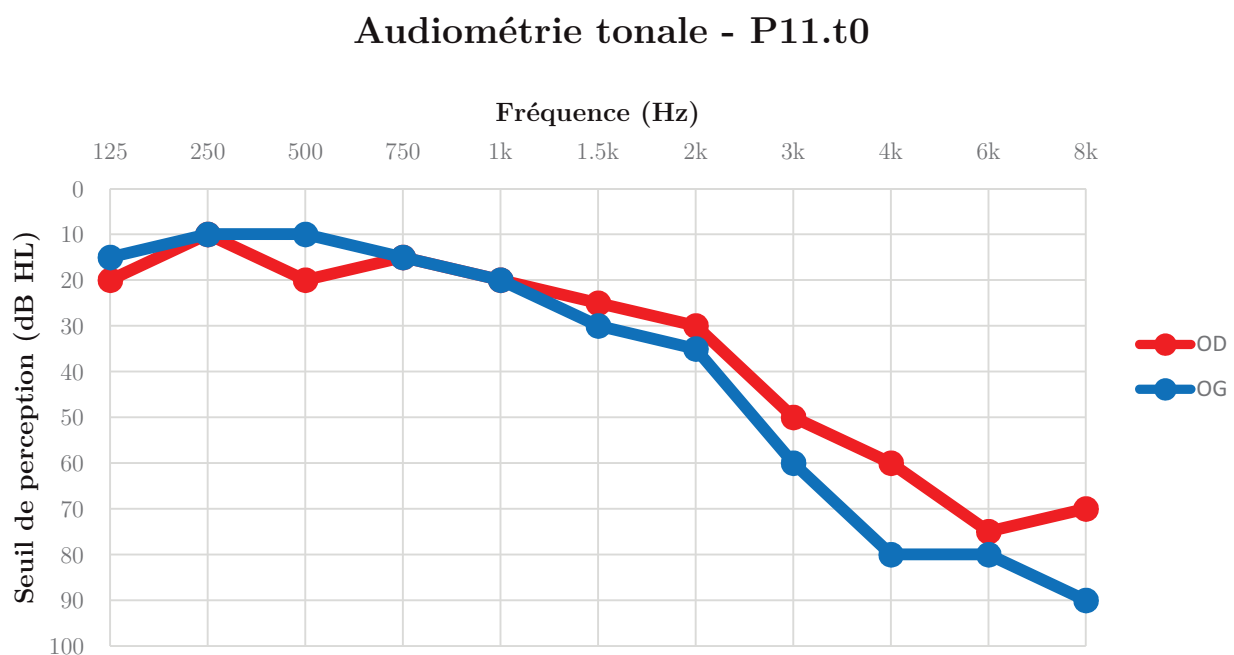


Figure 3.1. Audiométrie tonale d'un patient presbyacousique de l'étude 4. Dans la légende, OD signifie *oreille droite*, et OG *oreille gauche*.

L'audiométrie vocale donne un indice d'audibilité des sons de parole, en quantifiant le taux d'erreur dans la répétition de mots pour des intensités de stimulation décroissantes. Différentes listes de matériel phonétique existent (test dissyllabique de Fournier, liste phonétique de Lafon), qui peuvent être dotés de sens ou non (logatomes). La compréhension de la parole peut aussi être testée dans le bruit (logatomes de Dodelé), en faisant varier le

rapport signal-sur-bruit avec des bruits de différentes natures (conversation type cocktail party, bruit blanc ou coloré, *et cetera*).

Les mesures suivantes ont davantage rapport à des mécanismes cognitifs que sensoriels ou perceptifs, et informent donc surtout sur des déficits centraux liés à l'audition.

Les tests d'écoute dichotique testent les capacités attentionnelles, en présentant des sons différents à chaque oreille, et demandant au sujet de restituer certaines informations : comme il est très ardu d'assimiler l'intégralité des informations de deux sources concurrentes, il sera possible de quantifier la provenance des informations restitués pour mettre en évidence une plus grande attention portée sur une oreille par rapport à l'autre ; ou, en demandant au sujet de ne faire attention qu'à une latéralité, il est possible de compter le nombre d'erreurs et donc de chiffrer les performances attentionnelles auditives.

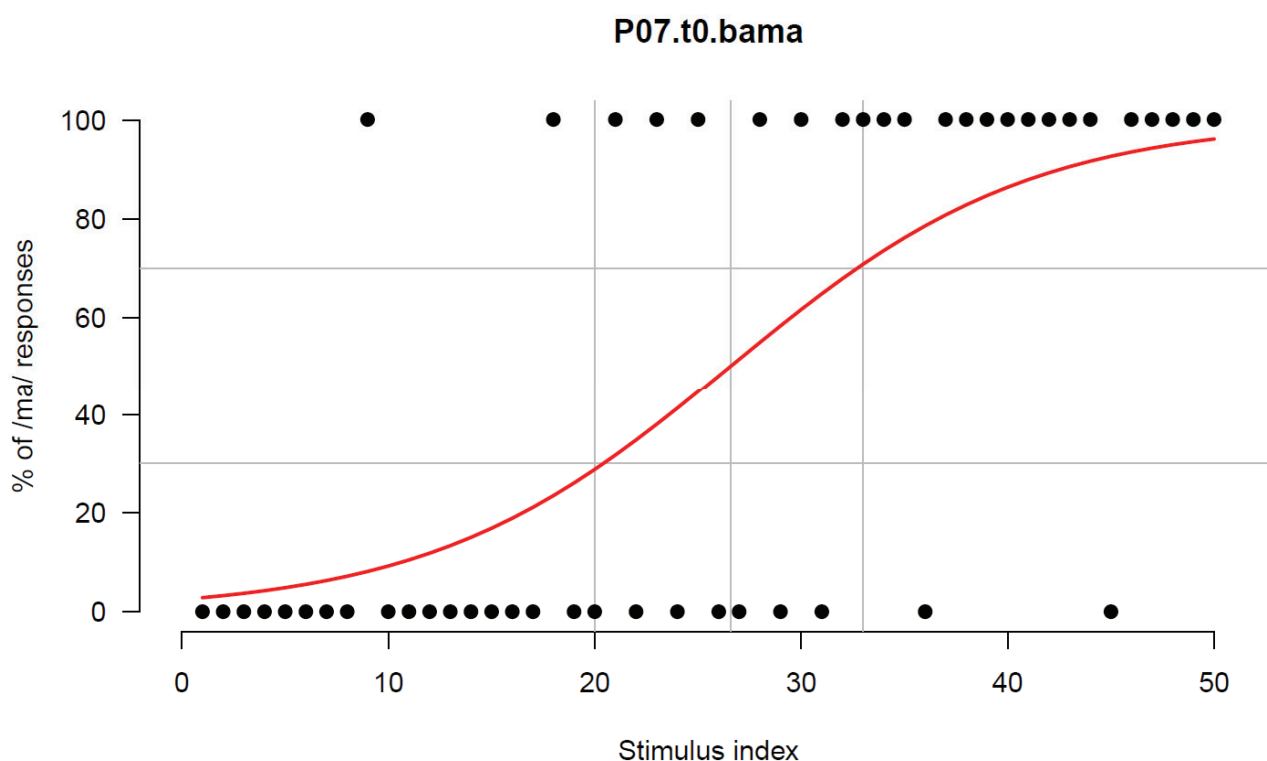


Figure 3.2. Courbe d'identification phonémique d'un patient presbycousique de l'étude 4, pour un *continuum* de nasalité /ba/-/ma/. Les réponses d'identification sont représentées par les points noirs, pour chacun des 50 *stimuli* présentés (point en bas quand réponse /ba/, et en haut quand réponse /ma/). Elles permettent d'obtenir la courbe sigmoïde, en rouge.

Le test de perception catégorielle comprend les parties d'identification phonémique, et de discrimination phonémique. Ils permettent d'estimer la faculté à grouper les *stimuli* d'un *continuum*, sorte de dégradé acoustique, en des catégories distinctes, procédant ainsi à une extraction d'information invariante à partir de *stimuli* physiques variables. Dans la partie

d'identification, les *stimuli* d'un *continuum* sont présentés un par un, et le sujet doit indiquer la syllabe perçue après chaque présentation (Figure 3.2., points noirs). Le test de discrimination existe sous différents formats, l'un des plus connus consistant à faire écouter consécutivement les deux *stimuli* d'une paire, suite à quoi le sujet devra répondre s'il a perçu les syllabes comme identiques ou différentes. Si l'analyse des scores d'identification résulte en une courbe en S (dite sigmoïde ; Figure 3.2., en rouge), et que la discrimination est plus importante pour les paires de *stimuli* situés de part et d'autre de la frontière catégorielle que pour les paires comprises dans une même catégorie, alors la perception est jugée catégorielle.

3.1.2.2. Anatomo-physiologie

Sur le plan anatomo-physiologique, soit de la structure ou du fonctionnement du système auditif humain sans rapport à l'impression du sujet, les examens cliniques sont nombreux et apportent un éclairage sur des étapes distinctes de la perception de parole.

L'examen physique est un premier pas nécessaire, permettant de détecter les anomalies morphologiques évidentes telles que l'aplasie. L'otoscopie permet ensuite de s'assurer de l'absence d'un bouchon de cérumen dans le conduit auditif externe, ainsi que de la bonne structure et apparence de la membrane tympanique. La tympanométrie mesure la réactivité du tympan en termes de souplesse, suite à l'application d'une variation de pression dans le conduit auditif externe. Elle mesure également l'existence et le bon fonctionnement du réflexe stapédien. Toutes ces techniques sont aptes à mettre en évidence une éventuelle surdité de transmission.

Au niveau de la cochlée, l'enregistrement des oto-émissions acoustiques spontanées ou évoquées par un son fournit un indice de l'intégrité fonctionnelle et anatomique des cellules ciliées externes de la cochlée. Lors de leur activité contractile, ces cellules sensorielles émettent un son, qui peut être capté par une sonde microphonique placée dans le conduit auditif externe. L'oreille ne perçoit donc pas simplement les sons : elle en émet aussi, qui permettent alors d'estimer la fonctionnalité du système efférent olivo-cochléaire médian, régulateur de l'activité des CCE. En pratique, cette fonctionnalité s'examine en administrant un bruit blanc dans l'oreille controlatérale à celle où sont enregistrées les oto-émissions acoustiques, et en observant une diminution d'amplitude de ces dernières.

3.1.3. Electrophysiologie de l'audition

Nous allons désormais entrer dans le domaine des potentiels évoqués. Comme dans tout jargon technique, l'adjectif « évoqué » a ici un sens bien particulier : il désigne une activité dont la latence est toujours identique en réponse aux répétitions d'un même *stimulus*. Il s'oppose à « induit » qui désigne une réponse au *stimulus* qui aura une latence variable au fur

et à mesure des répétitions, comme cela peut être le cas à des niveaux de traitement particulièrement élevés, en proie à des mécanismes attentionnels fluctuant de façon incontrôlable pour l'expérimentateur.

Les premiers enregistrements EEG chez l'Homme remontent aux travaux de Hans Berger, en 1924. Quelques années plus tard, Wever & Bray découvrent chez le chat ce qu'ils croient être une activité du nerf auditif, mais qui se révélera quelques années plus tard être une activité cochléaire (Wever & Bray, 1930 ; Davis *et al.*, 1934).

L'*électrocochléographie* ainsi nommée est l'enregistrement de l'activité électrique cochléaire générée par les influx ioniques dans les cellules ciliées internes et externes. Elle différencie trois types d'activités : le *potentiel microphonique cochléaire*, le *potentiel de sommation* et le *potentiel d'action composé*, ce dernier étant généré par le nerf auditif (Eggermont, 1974 ; Eggermont & Odenthal, 1974). Les potentiels cochléaires microphoniques révèlent par ailleurs la capacité des cellules ciliées à encoder temporellement par suivi de phase les fréquences jusqu'à environ 5000 Hz : chaque mouvement vers le haut de la membrane basilaire entraîne un flux ionique entrant dans les cellules ciliées du fait de la contrainte mécanique appliquée sur leurs stéréocils ; un mouvement vers le bas n'entraînera par conséquent pas de réponse des cellules ciliées, qui ne transmettent donc qu'un seul message par cycle du *stimulus* acoustique.

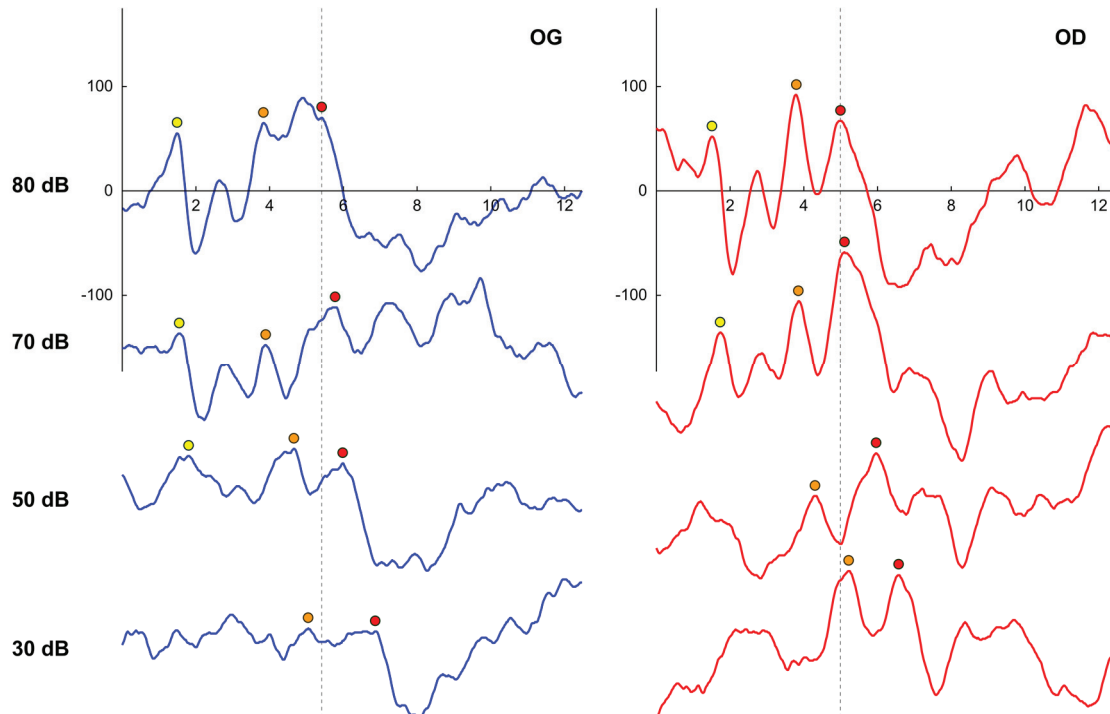


Figure 3.3. Potentiels Evoqués Auditifs Précoces (PEAP) en réponse à des clics, chez un sujet normo-entendant (données acquises durant la mise en place du protocole de l'étude 4). Les ondes I, III et V sont représentées par les points jaunes, orange et rouges, respectivement. L'onde II est également observable à 80 et 70 dB nHL. Notons l'effet de l'intensité de stimulation sur la latence de l'onde V.

Les *potentiels évoqués auditifs précoces* (PEAP), ou réponses du tronc cérébral auditif aux clics (en anglais Click ABR pour *Auditory Brainstem Response*), sont une série de 7 ondes arrivant dans les 10 millisecondes suivant l'écoute d'un clic (Jewett & Williston, 1971). Ce *stimulus* suit théoriquement une distribution de Dirac, c'est-à-dire qu'elle vaut l'infini en zéro, et zéro partout ailleurs ; autrement dit, un clic est censé contenir toutes les fréquences en une durée infiniment petite. Ils sont enregistrés à l'aide d'un différentiel d'électrodes placées au vertex (sommet du crâne), et à la mastoïde ou au lobe du côté de l'oreille stimulée. Les latences des différentes ondes informent sur l'intégrité des voies auditives ascendantes sous-corticales, du nerf auditif au colliculus inférieur, comme nous allons le voir en détail dans la prochaine partie. Aussi, la détection de l'onde V, la plus robuste à intensité de stimulation basse (Figure 3.3.), informe de l'audibilité des hautes fréquences de 2 à 4 kHz (Stapells & Oates, 1997).

Du fait que le clic contienne théoriquement toutes les fréquences, l'ensemble de la membrane basilaire (plus précisément des cellules ciliées internes qui s'y situent) va transmettre le signal au nerf auditif. Or nous avons vu que les hautes fréquences sont encodées sur la membrane dans la partie basale de la cochlée, et les basses fréquences à l'apex ; en « entrant » par la fenêtre ovale, les hautes fréquences vont atteindre leur cible avant les moyennes, et encore plus avant les basses fréquences, si bien que le nerf auditif recevra une séquence d'afférences des différentes fréquences. Comme l'amplitude des pics des ABR dépend grandement de la taille de la population neuronale synchronisée, l'idée est apparue de réarranger temporellement les différentes composantes spectrales du clic pour qu'elles arrivent toutes au même instant aux cellules ciliées puis au nerf auditif. Cette technique appelée *Chirp ABR* permet alors d'obtenir des réponses plus amples et plus facilement détectables (Dau *et al.*, 2000).

La réponse de suivi de fréquence, ou *Frequency-Following Response* (FFR) est une activité évoquée soutenue qui représente un *stimulus* périodique en s'y alignant en phase (Worden & Marsh, 1968 ; Marsh *et al.*, 1970). Elle s'observe entre 6 et 20 ms après le début du *stimulus*, qui peut être un ton pur d'apparence sinusoïdale lorsqu'il ne contient qu'une seule fréquence, ou un ton complexe lorsqu'il possède plusieurs fréquences (ou encore un son de parole, comme nous allons le voir dans la partie suivante. Cette propriété de suivi de phase (*phase-locking*) donne à la cette réponse l'apparence du *stimulus* qui l'évoque.

Les réponses stationnaires auditives, ou *Auditory Steady-State Response* (ASSR) diffèrent de la FFR en suivant la fréquence (temporelle) de modulation en amplitude d'une fréquence (spectrale) porteuse, plutôt que la fréquence porteuse même (Galambos *et al.*, 1981). Cette réponse est très utilisée dans l'implémentation de mesures objectives de l'audition, car elle apporte une information semblable à l'audiométrie tonale sans nécessiter la participation active du sujet (Herdman & Stapells, 2003). Elles s'enregistrent comme les autres ABR.

En continuant de monter dans les voies auditives nerveuses humaines, nous arrivons aux *Potentiels Evoqués Auditifs de latence moyenne*, avec les ondes Pa, Na, Pb et Nb arrivant à des latences comprises entre 20 et 40 ms après la stimulation. La lettre indique le signe de l'onde (Positive ou Négative).

Enfin, les Potentiels Evoqués Auditifs corticaux forment un complexe de 3 ondes principales, P50, N1 et P2 arrivant 50, 100 et 200 ms après une variation conséquente du niveau d'énergie acoustique dans le *stimulus* (le nombre indique en effet la latence d'apparition de l'onde ; Figure 3.4.).

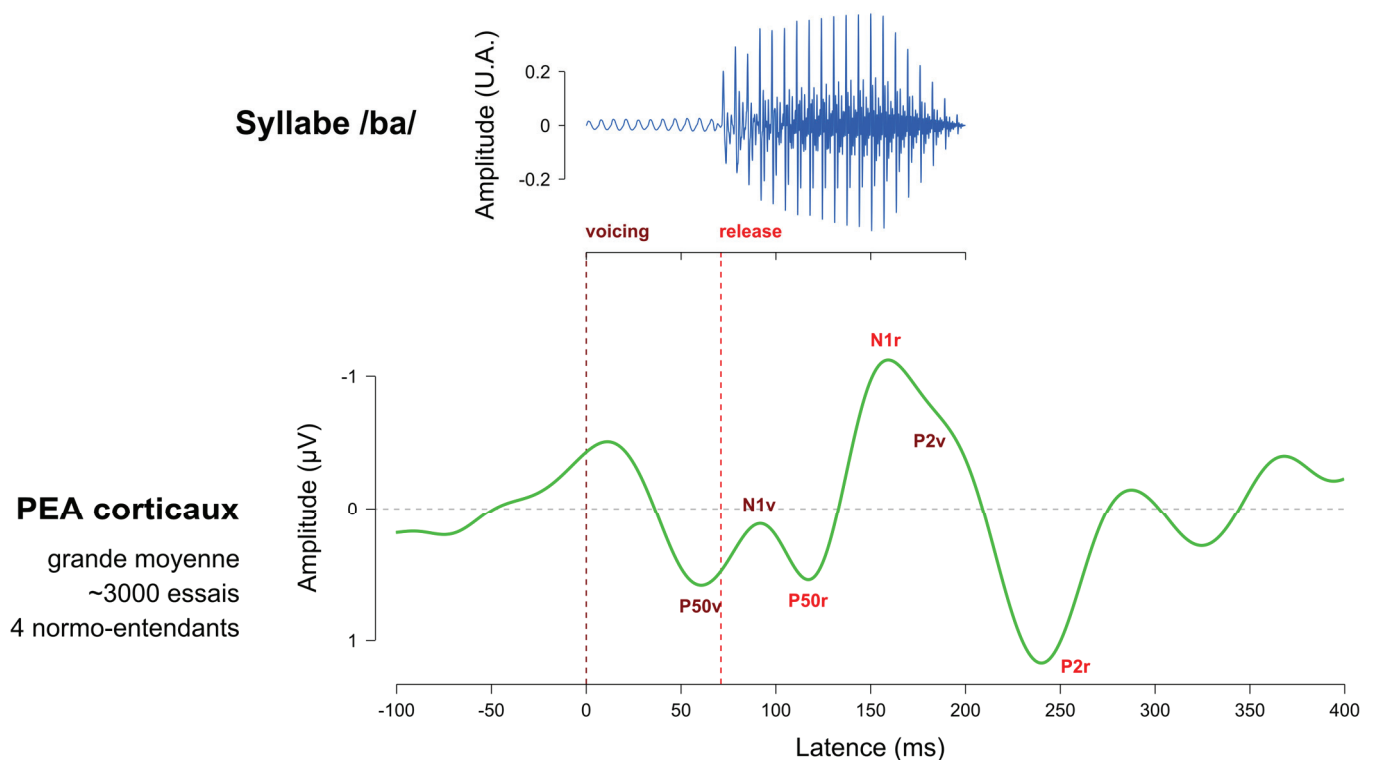


Figure 3.4. Grande moyenne des Potentiels Evoqués Auditifs corticaux de 4 participants normo-entendants en réponse à une syllabe /ba/, observée en Cz référencée au nez (figure réalisée à l'aide des données de l'étude 4). Un premier complexe P50-N1-P2 s'observe en réponse au début du *stimulus* (voisement, ou *voicing* ; en bordeaux), et un second est évoqué par l'explosion du /b/ (relâchement de l'occlusion, ou *release* ; en rouge).

D'autres ondes corticales liées à l'audition s'observent avec l'EEG, telles que la *Mismatch Negativity* (MMN) ou la P300. La MMN est une réponse à l'occurrence d'un intrus dans une séquence régulière de sons, consistant donc en quelque sorte à une erreur de prédiction codant la surprise dans le cerveau auditif. La P300 se manifeste pour un *stimulus* attendu, et traduit donc un effet de l'attention.

D'autres outils que l'EEG sont disponibles pour étudier le système auditif humain. La MEG apporte une information complémentaire à l'EEG, le champ magnétique étant

orthogonal au champ électrique, et décroissant plus rapidement avec la distance. Des PEAP ont tout de même pu être enregistrés malgré cette dernière propriété, avec l'avantage donné par la MEG d'un grand nombre de capteurs permettant une reconstruction de source plus efficace qu'en EEG de scalp (Parkkonen *et al.*, 2009).

Comme nous l'avons vu dans la partie précédente au niveau de la description du parcours cortical de l'information de parole, l'IRMf est utile pour localiser les aires cérébrales impliquées dans un traitement particulier, pour la mise en évidence duquel le protocole expérimental est conçu. Aussi, elle permet de quantifier l'activation d'une aire par contraste à une autre, à la même aire dans une autre condition expérimentale, ou encore à la même aire chez un sujet appartenant à un groupe différent. Elle met cependant plusieurs secondes à mesurer le signal BOLD (*Blood-Oxygen-Level Dependent*) pour l'ensemble du volume cérébral, et ne permet donc pas de suivre dans le détail l'encodage des sons de parole.

3.2. Speech ABR

La réponse du tronc cérébral auditif aux sons de parole (Speech ABR pour *Auditory Brainstem Response*, bien qu'elle soit désignée parfois par d'autres étiquettes telles que speech-evoked ABR, complex ABR, ou FFR to speech) est une activité électrophysiologique évoquée reproduisant finement et fidèlement un grand nombre de détails spectro-temporels du *stimulus* de parole (Russo *et al.*, 2004 ; Johnson *et al.*, 2005).

Elle se compose de deux types de composantes : transitoires (*onset* et *offset response*) et soutenue (FFR ; Figure 3.5.). Les premières sont évoquées par des changements brefs du niveau d'énergie acoustique dans le *stimulus* : une augmentation entraîne une réponse *onset*, une diminution une réponse *offset*. La composante FFR présente la même propriété de suivi de fréquence que son homonyme évoquée par des tons purs ou complexes, dont elle ne changerait qu'au niveau du *stimulus* utilisé, ici un son de parole.

Les premiers enregistrements de cette réponse datent du milieu des années 1990, mais elle n'a commencé à être intensément étudiée qu'au milieu des années 2000. Aussi, relativement peu d'équipes utilisent ce marqueur, ce qui polarise le paysage bibliographique et ancre les pratiques méthodologiques. Les principaux chercheurs ayant fait avancer nos connaissances sur cette réponse sont Nina Kraus, Gary Galbraith et Ananthanaryan Krishnan, chacun ayant exploré de leur côté des voies relativement distinctes, avec des variations méthodologiques propres.

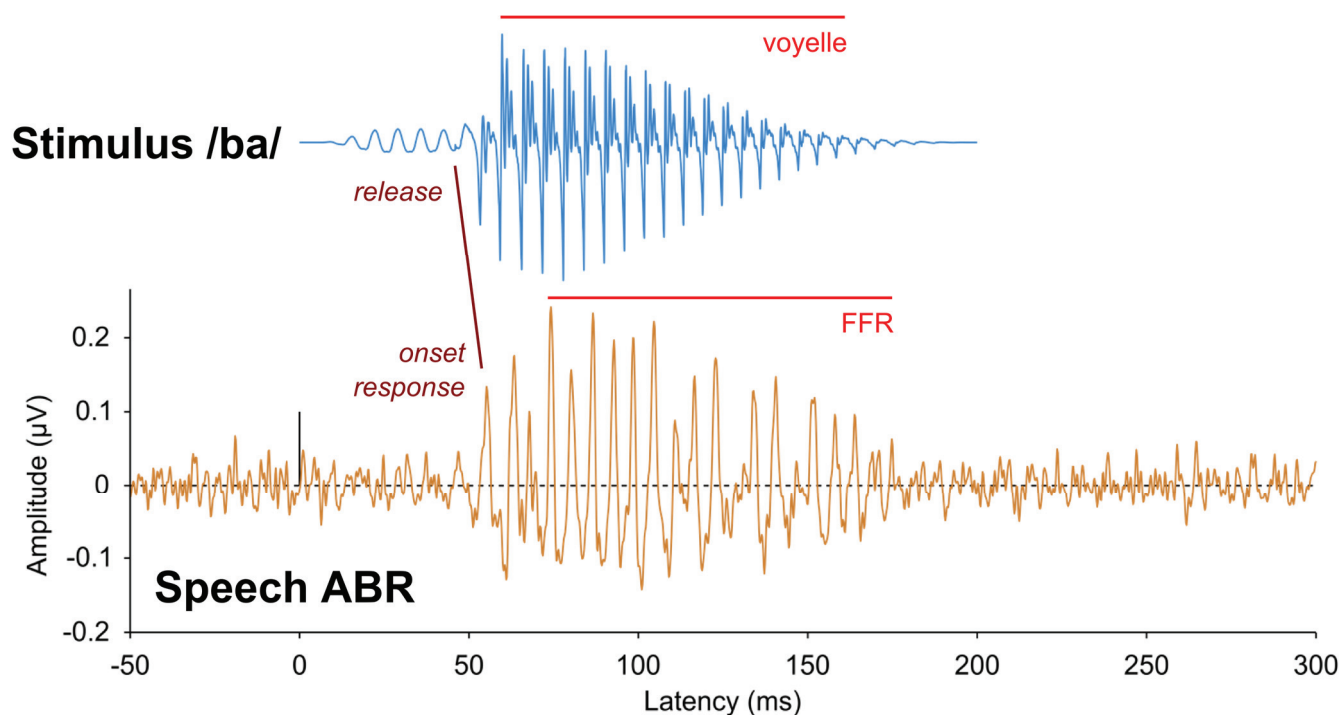


Figure 3.5. Grande moyenne des Speech ABR évoquées par une syllabe /ba/, observée en Cz reréférencée à la moyenne des mastoïdes (11 participants, pour un total de près de 22300 essais ; données issues de l'étude 2). Nous observons bien une grande ressemblance entre la réponse (en bas, courbe orange) et le *stimulus* (en haut, courbe bleue), ainsi que la présence des deux types de composantes de la Speech ABR, avec l'*onset response* et la FFR.

Nous allons maintenant explorer la méthodologie de la Speech ABR, en partant de son protocole de stimulation et de recueil pour arriver aux différentes stratégies d'analyse de données ; nous passerons ensuite en revue la littérature des générateurs neuronaux de cette réponse qui, comme son nom l'indique, est censée prendre source au niveau sous-cortical ; puis nous aborderons les différents champs d'application de ce marqueur électrophysiologique de l'encodage des sons de parole, pour enfin les synthétiser en un état des lieux de nos connaissances actuelles sur ce marqueur électrophysiologique de notoriété croissante.

3.2.1. Méthodologie

En 2010 a été publié un tutoriel très complet sur le recueil de la Speech ABR, présentant les différentes variations méthodologiques rencontrées dans la littérature (Skoe & Kraus, 2010). Nous allons ici développer certains points plus en détail, et en actualiser d'autres à la lumière des études parues depuis.

3.2.1.1. Stimulation

Stimulus

Depuis le début des enregistrements de Speech ABR, le matériel de stimulation a connu de nombreuses variations toutefois restreintes autour d'un petit groupe de syllabes. L'équipe de Nina Kraus, grâce à qui nous devons l'augmentation massive de la littérature sur la Speech ABR, a très largement utilisé une syllabe /da/ synthétique d'une durée de 40 ms, avec une F0 variant entre 103 et 125 Hz (King *et al.*, 2002 ; Russo *et al.*, 2004 ; Figure 3.6.). Celle-ci est à présent largement diffusée, et utilisée par de nombreuses équipes débutant dans l'utilisation de ce marqueur électrophysiologique (Sinha & Basavaraj, 2010 ; Filippini *et al.*, 2012 ; Gnanateja *et al.*, 2012 ; Elkabariti *et al.*, 2014 ; Malayeri *et al.*, 2014 ; Jafari *et al.*, 2014, 2015). Nous allons cependant citer les autres *stimuli* existant, variant dans leur identité, leur durée, leur composition spectrale et leur naturalité.

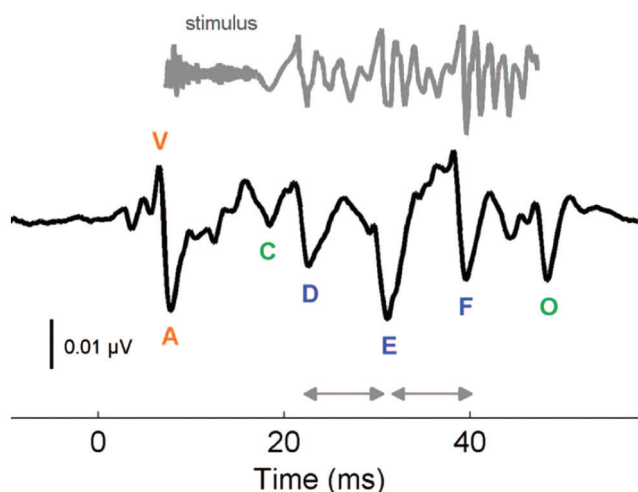


Figure 3.6. Syllabe /da/ synthétique de 40 ms (en haut) et Speech ABR correspondante (en bas ; Skoe & Kraus, 2010a). Le *stimulus* est décalé vers la droite de 7 ms pour rendre plus directe sa comparaison avec la Speech ABR. Les pics V et A correspondent à l'*onset response*, les pics C et O à des réponses transitoires (O étant l'*offset*) car évoqués par des parties transitoires du *stimulus*, et les pics D, E et F correspondent à la FFR. Enfin, les flèches grises horizontales représentent la période du *stimulus*, soit 110 Hz en moyenne (période d'environ 9 ms).

● Sons de parole utilisés

Dans la grande majorité des études s'intéressant à la Speech ABR, les sons de parole utilisés ont été des syllabes consonne-voyelle, avec une consonne occlusive voisée comme /d/ ou /b/ (Richard *et al.*, 2010). Ce choix est justifié par l'énergie acoustique importante libérée lors de la désocclusion, qui s'apparente à un clic et évoque par conséquent une composante *onset* très remarquable (Skoe & Kraus, 2010a). Le choix de la voyelle se porte très souvent sur /a/, bien que /i/ et /u/ évoquent des FFR plus amples (Choi *et al.*, 2013). Cela s'explique par la valeur du premier formant F1 très basse (autour de 300 Hz) de ces deux derniers phonèmes : le pic évoqué par F1 dans la FFR empièterait alors sur le pic évoqué par la F0 (entre 100 et 200 Hz le plus souvent). La voyelle /a/ possède un F1 plus haut, autour de 700 Hz, évitant de ce fait une telle superposition F1 - F0.

Afin d'étudier le suivi des variations de F0 (lié au pitch), des syllabes de langues tonales comme le mandarin sont aussi utilisées (Krishnan *et al.*, 2004 ; Song *et al.*, 2008 ; Jeng *et al.*, 2011d). Exceptionnellement, nous retrouvons des mots (Kouni *et al.*, 2014), voire des groupes de mots (Galbraith *et al.*, 2004 ; Choi *et al.*, 2013). La principale raison pour laquelle les sons de parole étudiés sont principalement restreints à la syllabe est le nombre requis d'essais, c'est-à-dire de fois où le *stimulus* est présenté au sujet, pour obtenir une réponse émergeant du bruit en un temps donné. Nous allons voir ce point plus en détail dans la suite du présent chapitre.

- Durée

Avec pour minimum les 40 ms de la syllabe /da/ la plus répandue, elle peut monter jusqu'à 500 ms pour des mots, mais est souvent fixée aux alentours de 200 ms. Le choix de la durée est primordial, à la fois pour contrôler le temps d'acquisition car plus le *stimulus* est long et plus cela prendra de temps à enregistrer un nombre donné d'essais, mais aussi pour obtenir suffisamment de cycles dans la FFR et ainsi pouvoir effectuer des analyses plus complètes (Skoe & Kraus, 2010).

- Composition spectrale

La F0 détermine à la fois la fréquence de la FFR, mais aussi la résolution de ses formants. Par exemple, pour une F0 de 100 Hz, une F1 située autour de 800 Hz et d'une largeur de bande de 200 Hz de part et d'autre sera dessinée par 5 harmoniques (600, 700, 800, 900 et 1000 Hz) ; pour une F0 de 170 Hz, le même formant n'englobera plus que 2 harmoniques (680 et 850 Hz). La fameuse syllabe /da/ de 40 ms a une F0 variant progressivement entre 103 et 125 Hz, donc située autour de 111 Hz.

- Naturalité

La synthèse de parole est souvent utilisée dans les études de Speech ABR, étant donné qu'il est tentant de contrôler finement tous les paramètres spectro-temporels de la syllabe désirée, qui régiront ensuite la forme d'onde et le contenu spectral sous-jacent de la Speech ABR. Si le recours aux techniques de synthèse est indispensable dans les études désirant comparer plusieurs syllabes, la parole naturelle contient plus d'énergie dans les hautes fréquences, et véhicule une information plus riche (identité du locuteur, état émotionnel, entre autres). Une étude récente montre la sensibilité de la Speech ABR à un contexte enrichi par les voix naturelles de différents locuteurs, en comparaison de la réponse évoquée par la voix d'un seul locuteur ; cette sensibilité se manifeste par un encodage accru des indices spectraux dans la FFR (Nan *et al.*, 2015).

En élargissant la notion de naturalité du *stimulus* à sa qualité linguistique et complexe plutôt qu'à une simple opposition aux phonèmes de synthèse, plusieurs faisceaux de preuves ont mis en évidence un encodage préférentiel des sons de parole, lorsqu'ils sont porteurs d'indices acoustiques connus. Par exemple, l'amplitude de la FFR augmente lorsque le signal de parole est présenté à l'endroit, donc intelligible (Galbraith *et al.*, 2004). Aussi, une dissociation de l'effet du bruit sur les Click ABR et Speech ABR, en l'occurrence ne dégradant que la réponse aux sons de paroles, plaide pour un mécanisme spécifique ou au moins préférentiel d'encodage des sons complexes (Song *et al.*, 2006).

Stimulateur

L'artéfact de stimulation est la présence du *stimulus* de parole dans les données électrophysiologiques, dû au rayonnement électromagnétique des écouteurs sur les électrodes de recueil du signal EEG. Lorsqu'il arrive pour les PEAP, cela n'importe que peu étant donné que le clic dure 100 μ s environ, et que la première réponse des PEAP, soit l'onde I, arrive peu après 1 ms : il n'y a pas superposition du signal d'intérêt et de l'artéfact de stimulation.

La Speech ABR est évoquée par un son de parole pouvant durer entre 40 et 400 ms, et ses différentes composantes arrivent à des latences comprises entre 6 et 20 ms. Par conséquent, un artéfact de stimulation potentiel se superposerait au signal d'intérêt, le rendant inutilisable d'autant plus du fait de la grande ressemblance spectro-temporelle entre Speech ABR et *stimulus*.

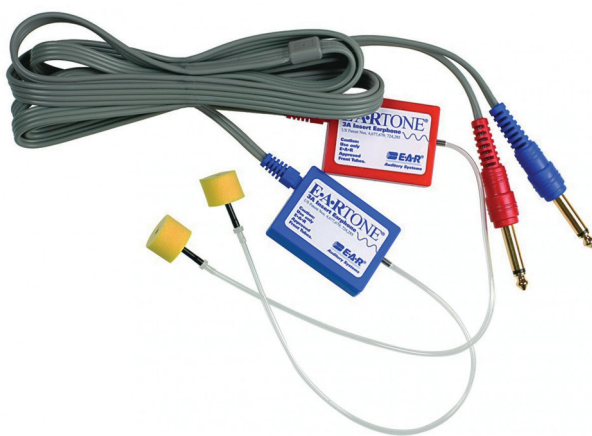


Figure 3.7. Exemple d'écouteurs de type *inserts*.

Plusieurs techniques existent pour tenter de supprimer l'artéfact de stimulation lorsqu'il se retrouve dans les données, mais il reste souhaitable de faire en sorte qu'il ne s'y retrouve pas. Pour cela, des stimulateurs bien particuliers sont utilisés pour évoquer la Speech ABR : les écouteurs de type *inserts* (*insert earphones* ; Figure 3.7.), d'utilisation courante en audiométrie. Ils consistent en un transducteur électroacoustique prolongé par un tube souple de quelques dizaines de centimètres (généralement 30 cm), qui permet alors d'éloigner le transducteur des électrodes et de leurs câbles.

L'utilisation d'inserts permet souvent d'éviter la contamination des données par un artéfact de stimulation, bien qu'il puisse parfois polluer le signal quand l'impédance des électrodes est mauvaise (trop élevée, correspondant à une électrode mal collée et captant plus l'environnement électromagnétique que le signal électrophysiologique) ou quand l'intensité de stimulation est trop forte (l'amplitude du rayonnement électromagnétique polluant dépend logiquement de ce paramètre ; Akhoun *et al.*, 2008b). Néanmoins, les inserts ne sont pas neutres sur l'acoustique du signal envoyé : leurs tubes en matière plastique entraînent une distorsion du son à partir de 1 kHz, et une forte atténuation des fréquences après 3 à 4 kHz. L'intensité de stimulation maximale est également limitée, ne pouvant dépasser 110 dB SPL sans subir d'importantes distorsions. Cette dernière raison explique peut-être le peu d'études effectuées chez les malentendants.

Paramètres de stimulation

- Cadence de stimulation

La cadence de stimulation se définit comme le nombre de *stimuli* délivrés par seconde. Elle peut être décrite par l'intervalle inter-*stimulus* (ISI), représentant la durée entre la fin du *stimulus* n et le début du *stimulus* $n+1$, ou par le paramètre de *Sound Onset Asynchrony* (SOA), représentant l'écart temporel entre le début de deux *stimuli* consécutifs, soit la somme entre durée du *stimulus* et l'ISI. Par exemple, dans une expérience où des syllabes de 200 ms sont espacées d'un intervalle fixe de 50 ms (le précité ISI), le SOA sera de $200 + 50 = 250$ ms, et la cadence de stimulation de $1000 / 250 = 4$ *stimuli* par seconde.

De la même manière que la durée du *stimulus*, la cadence de stimulation impacte la durée totale d'enregistrement. Augmenter ce rythme est donc tentant, mais des effets ont été montrés sur les deux composantes de la Speech ABR : une cadence plus soutenue entraîne un allongement de la latence de l'*onset response*, ainsi qu'une baisse d'énergie au-delà de 400 Hz dans le contenu spectral de la FFR (Krizman *et al.*, 2010). Notons au passage que la latence des réponses aux clics ne change pas en fonction de ce facteur, alors que l'amplitude et la latence des ondes corticales sont modifiées : la Speech ABR a donc sur ce point un comportement plus proche de celui des PEA corticaux.

L'ISI peut être fixe (50 ms dans l'exemple précédent) ou variable (50 ms plus ou moins 20 ms), auquel cas l'intervalle de variation est appelé *jitter*, et permet de rendre moins prévisible l'arrivée du *stimulus* suivant, en cassant la régularité du rythme de délivrance des *stimuli*. Les effets de ce *jitter* sur la Speech ABR n'ont toutefois pas été établis.

Enfin, notons la proposition récente d'un mode de présentation des *stimuli* composé de courts et de longs ISI, permettant d'acquérir la Speech ABR et les PEA corticaux dans un même bloc tout en garantissant leur qualité (Bidelman, 2015a).

- Latéralité de stimulation

La présentation du *stimulus* à une seule oreille (stimulation monaurale) permet de mettre en évidence une asymétrie dans le traitement de la parole par les voies nerveuses auditives : chez des adultes droitiers normo-entendants, il a été montré que la Speech ABR à une syllabe présentée en monaural à l'oreille droite, comparativement à l'oreille gauche, avait une FFR et une *offset response* avancées en latence, ainsi qu'un encodage plus robuste pour les composantes spectrales au-delà de F0, notamment au niveau du premier formant F1 ; la latence de l'*onset response* et les écarts inter-pics dans la FFR n'étaient cependant pas significativement différents entre les conditions de latéralité de stimulation (Hornickel *et al.*, 2009a). Ici encore, les Click ABR ne montraient pas d'effet de latéralité.

La stimulation binaurale permet quant à elle d'évoquer des réponses plus amples et plus robustes (Ahadi *et al.*, 2014b), sans doute car bénéficiant de l'effet d'amplification binaurale (Binaural Interaction Component, environ + 6 dB), manifestation supplémentaire de non-linéarité dans les voies auditives sous-corticales, et mise en évidence dans la littérature des Click ABR et de la FFR (Dobie & Berlin, 1979). Le choix de la latéralité de stimulation doit donc se baser sur l'hypothèse à tester et son lien à l'asymétrie du système auditif humain.

- Niveau sonore

La plupart des études testant des sujets jeunes et normo-entendants fixe le niveau sonore de stimulation à une intensité de 80 dB SPL. Une étude antérieure a montré que l'amplitude et la latence de l'*onset response* et de la FFR variaient en fonction de l'intensité de stimulation : lorsque le niveau sonore diminue, l'amplitude de ces composantes diminue et leur latence se rallonge, d'ailleurs de façon plus importante pour la FFR (Akhoun *et al.*, 2008a). Cependant, une intensité de stimulation forte sera plus difficilement supportable pour le sujet, d'autant plus si la stimulation est binaurale, et augmentera la probabilité de déclencher le réflexe stapédien, l'artéfact physiologique du muscle post-auriculaire (voir plus loin), ou la contamination par l'artéfact de stimulation (Akhoun *et al.*, 2008b).

Du fait de l'influence de l'intensité de stimulation sur la latence des composantes de la Speech ABR, une attention particulière doit être portée à ce paramètre dans les études visant à analyser les moyennes de second niveau (ou grandes moyennes, à travers les participants). En effet, une intensité de stimulation mal contrôlée entraînera possiblement un déphasage entre les FFR des différents sujets, et donc une réponse moyenne dégradée (Figure 3.8.). L'intensité de stimulation peut être fixée dans deux référentiels : objectif (SPL, Sound Pressure Level, intensité physique) ou subjectif (SL, Sensation Level, intensité ressentie, par rapport au seuil de perception du sujet). Une question importante est donc de savoir quelle unité utiliser pour que les Speech ABR résultantes soient en phase à travers les sujets.

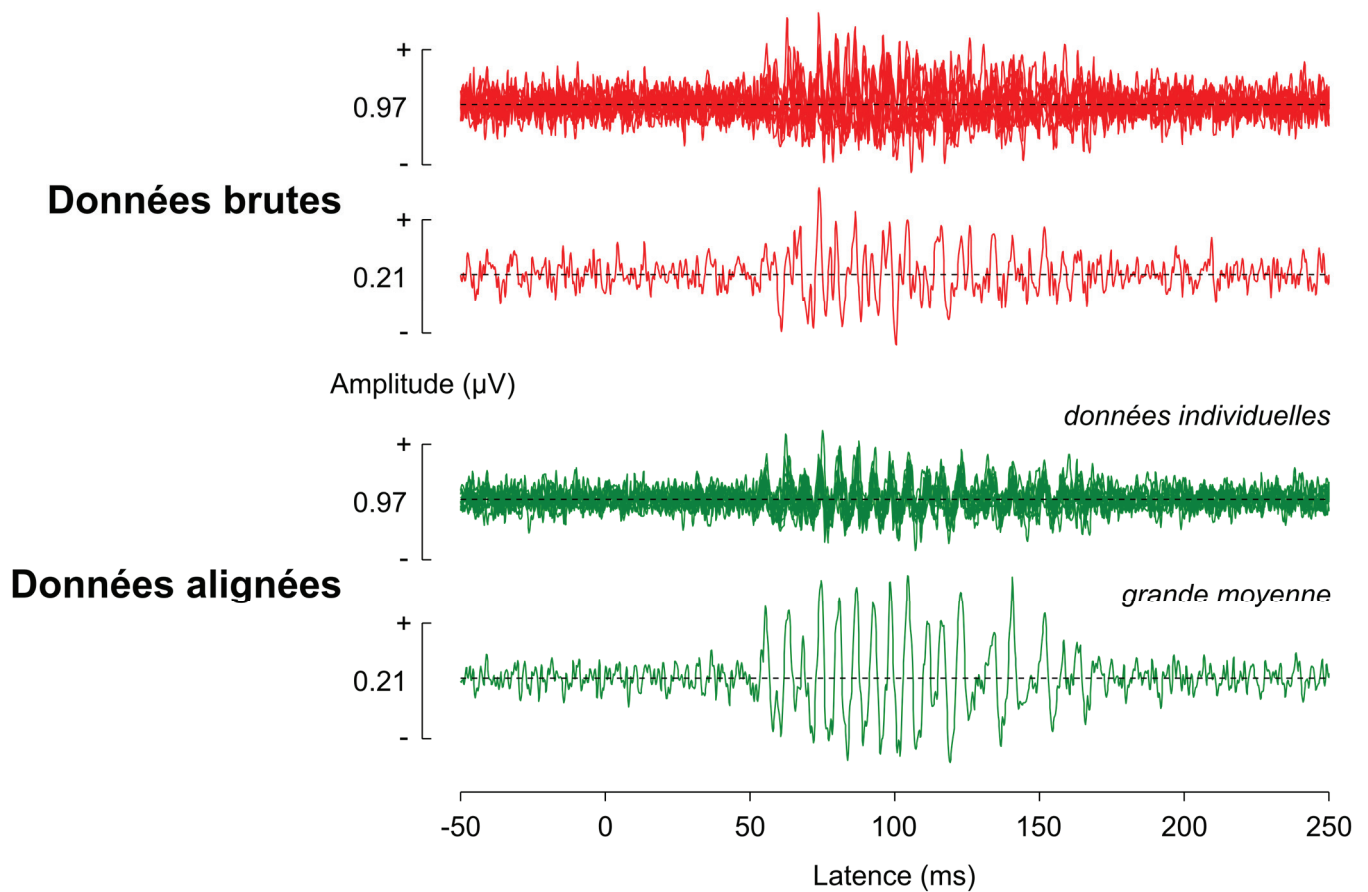


Figure 3.8. Effet délétère du déphasage des Speech ABR individuelles sur la qualité de la grande moyenne (courbes rouges, respectivement en haut et en bas), et efficacité de la stratégie de réalignement (courbes vertes). Les Speech ABR d'un groupe de 11 participants normo-entendants sont ici évoquées par la syllabe /ba/ représentée dans la Figure 3.5., et proviennent de l'étude 2 du présent manuscrit. Les données individuelles étaient déphasées en deux sous-groupes, du fait d'une modification matérielle du système de stimulation à mi-chemin du protocole. Cela s'observe notamment à l'endroit de l'*onset response*, à 55 ms : les données brutes moyennées présentent un double pic, en comparaison du simple pic plus ample dans la grande moyenne des données alignées.

En pratique, pour des participants normo-entendants, le choix entre dB SPL ou SL n'impacte que peu le résultat final, ceux-ci ayant par définition des seuils de perception proches, compris entre 0 et 20 dB HL (unité audiométrique ; 0 dB HL correspondant à environ 25 dB SPL) pour les fréquences pures de 250 à 8000 Hz, et ce pour les deux oreilles. En homogénéisant l'intensité de stimulation en dB SPL pour le groupe, les participants partageront des valeurs proches en dB SL.

Dans le cas des patients malentendants, nous avons observé qu'en fixant l'intensité en dB SL, la latence des FFR était parfaitement alignée entre les participants pourtant stimulés à des intensités physiques (SPL) variables. Un inconvénient de l'utilisation du dB SL est qu'il faut déterminer au préalable le seuil de perception du participant, ce qui prend nécessairement plus de temps qu'une calibration en dB SPL faite sur les transducteurs une seule fois avant utilisation pour tout le groupe.

- Polarité alternée

Pour dissocier l'activité neuronale des potentiels microphoniques cochléaires, engendrés par les flux ioniques entrant dans les cellules ciliées mécaniquement excitées, et pour réduire la contamination du signal physiologique d'intérêt par un artéfact de stimulation possiblement présent, la majorité des études de Speech ABR présente un même nombre de fois le *stimulus* dans une polarité et dans la polarité opposée (Campbell *et al.*, 2012). Théoriquement, l'artéfact de stimulation s'annulera, étant donné que la moyenne d'un signal et de son opposé est zéro ; par contre, la Speech ABR évoquée par le *stimulus* de chacune des polarités opposées contient une information indépendante de la polarité, davantage liée à l'enveloppe qu'aux détails spectraux fins. Ainsi, en faisant la moyenne des Speech ABR obtenues avec un *stimulus* et son opposé, nous obtenons la FFR d'enveloppe, tandis que la FFR spectrale s'obtient en moyennant la Speech ABR évoquée par le *stimulus* avec l'opposé de la Speech ABR évoquée par le *stimulus* de polarité opposée (Aiken & Picton, 2008). Plus simplement, en appelant S1 la réponse au *stimulus* de polarité positive, et S2 celle au *stimulus* de polarité négative, l'*envelope FFR* s'obtient par l'opération $(S1+S2)/2$, et la *spectral FFR* par $(S1-S2)/2$.

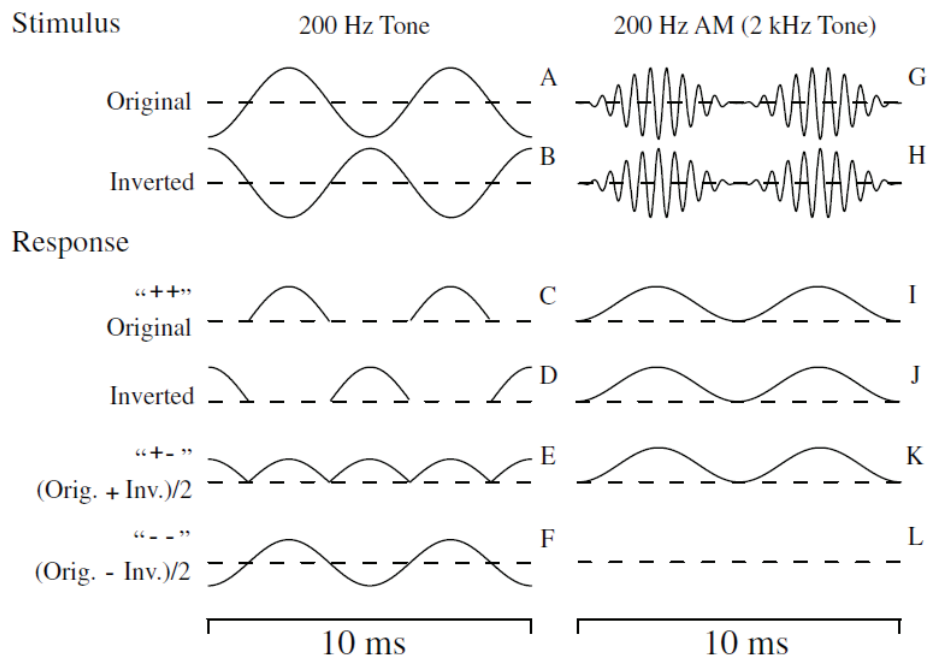


Figure 3.9. Schéma conceptuel de la polarité alternée (Aiken & Picton, 2008). Dans le code « ++ », « + - » et « - - », le premier signe informe de l'opération effectuée (+ pour somme, - pour soustraction), et le second des polarités du *stimulus* (+ pour identiques, - pour opposées). Dans le titre de la colonne de droite, AM signifie *amplitude modulated*. Le signal acoustique est encodé dans la cochlée pendant sa phase de raréfaction, c'est-à-dire lorsque la membrane basilaire se déplace vers l'intérieur du canal cochléaire. Dans la colonne de gauche, le ton pur de F0 200 Hz va donc être encodé en opposition de phase pour les polarités opposées (C et D) ; leur soustraction (F) donnera alors un signal neuronal proche du signal acoustique qui en est à l'origine : il s'agit de la FFR spectrale. Dans la colonne de droite, le nerf auditif suivra la modulation d'amplitude, avec une réponse identique aux *stimuli* de polarités opposées (I et J) ; leur moyenne (K) conservera cette information : c'est la FFR d'enveloppe.

Il est donc intéressant de disposer après l'enregistrement de l'information sur la polarité des *stimuli* délivrés, afin de pouvoir explorer ces deux versions de la FFR. Notamment, la FFR d'enveloppe représente davantage F0 et ses premières harmoniques, là où la FFR spectrale mettra en exergue les fréquences au-delà de 500 Hz, avec l'inconvénient de doubler l'artéfact de stimulation s'il s'avérait être présent dans les données.

Dans la pratique, les sons de parole utilisés sont bien rarement symétriques : les réponses aux polarités opposées seront ainsi différentes, et les FFR d'enveloppe et spectrale ne seront donc pas aussi pures que dans la Figure 3.9. Aussi, un effet de la polarité sur la Speech ABR a été montré dans une communication de congrès, et consiste en une altération en amplitude de l'encodage spectral hautes-fréquences pour une polarité par rapport à l'autre (*Aiken & Purcell, 2013*).

- Tâche et consignes au sujet

Dans la plupart des cas rencontrés dans la littérature, l'enregistrement des Speech ABR consiste en une tâche d'écoute passive. Le participant est installé dans un fauteuil conçu pour être confortable, et il lui est demandé de se détendre, d'éviter de trop bouger afin de ne générer qu'un minimum d'activité musculaire, mille fois plus ample que l'activité neuronale. Deux variantes s'observent alors : les patients peuvent regarder un film silencieux et sous-titré dans le cas d'une stimulation binaurale, voire même avec bande son diffusée à moins de 40 dB SPL dans le cas d'une stimulation monaurale où l'oreille non-stimulée reste disponible ; dans l'autre variante, il peut être demandé aux patients de se reposer, allant parfois jusqu'à les autoriser à fermer les yeux ou à dormir.

Pour certaines études, la participation active du sujet est requise pour répondre par exemple à une tâche d'écoute dichotique de détection de cible auditive (Lehman & Schönwiesner, 2014), ou pour déclarer un changement de percept (Galbraith *et al.*, 1997), entre autres.

3.2.1.2. Recueil

Système électroencéphalographique

La Speech ABR ayant dérivé de potentiels évoqués utilisés en clinique (Click ABR surtout, FFR dans une moindre mesure), elle partage avec ceux-ci de nombreux paramètres expérimentaux, dont le système d'enregistrement clinique. L'avantage principal de ce type de matériel est la fréquence d'échantillonnage élevée (souvent supérieure à 10 kHz), permettant donc d'observer l'activité électrophysiologique jusqu'à la moitié, voire le quart de cette fréquence, conformément au théorème d'échantillonnage de Nyquist-Shannon. Les systèmes EEG cliniques sont également très simples d'utilisation, avec des programmes de stimulation

« clé en main » permettant d'accéder directement à une Speech ABR filtrée et débarrassée des essais porteurs d'artéfacts

Cette simplicité constitue cependant un inconvénient pour les protocoles de recherche, en ce sens qu'elle rend difficile d'accès certains paramètres de stimulation. Il est ainsi moins possible de sophistiquer la cadence de stimulation, l'ordre des *stimuli* dans le cas du recueil de plusieurs Speech ABR en simultané, ou encore de définir un protocole actif avec recueil des réponses du participant. Au niveau du traitement des données, les essais unitaires ne sont parfois pas enregistrés, empêchant tout traitement de données post-acquisition tel que l'addition ou la soustraction des réponses aux *stimuli* de polarité alternée, ou le calcul de la consistance de la réponse à travers les essais. Enfin, les systèmes EEG cliniques permettent rarement l'enregistrement sur plus de deux ou quatre électrodes, ce qui ajoute une limite supplémentaire à l'exploration des différentes dimensions de la Speech ABR.

Type d'électrodes

Classiquement utilisées, les électrodes passives consistent en des pièces rondes et métalliques (argent/chlorure d'argent), qui captent le signal au niveau du scalp et le transmettent directement par voie électrique à l'amplificateur EEG. Les électrodes actives disposent quant à elles d'un préamplificateur du niveau de l'électrode même. La résistance au passage du courant entre le scalp et l'électrode est l'impédance : plus elle est basse, et plus les électrodes capteront efficacement le signal électrophysiologique (qu'il soit d'intérêt, ou considéré comme du bruit - activité musculaire dans le cas présent).

S'il est préférable de procéder aux enregistrements de Speech ABR dans une cabine insonorisée et isolée des interférences électromagnétiques de l'environnement (*faradisée*), l'utilisation d'électrodes actives permet de recueillir un signal de très bonne qualité dans des conditions moins propres, en amplifiant ce signal au niveau de l'électrode : un artéfact rayonnant sur le câble reliant l'électrode à l'amplificateur EEG aura donc un poids beaucoup moins important par rapport au signal.

Nombre et montage des électrodes

Par souci de simplicité (facilité et rapidité d'installation), la majeure partie des études de Speech ABR utilise 3 électrodes dans le cas d'une stimulation monaurale : l'électrode d'intérêt placée au vertex (Cz) ou plus en avant sur la ligne médiane (Fz voire Fpz), la référence placée à la mastoïde (TP9 à gauche ou TP10 à droite) ou au lobe d'oreille ipsilatéral à l'oreille stimulée, et la masse placée au front ou à la mastoïde contralatérale (sa position n'ayant pas d'influence sur le signal recueilli). Dans le cas d'une stimulation binaurale, une quatrième électrode est placée sur l'autre mastoïde ou lobe d'oreille (la masse est alors placée

au front), et la référence est alors définie à la moyenne des électrodes mastoïdiennes (ou des lobes d'oreille).

Ce placement des électrodes (vertex référencé aux mastoïdes) est appelé *montage vertical*, et permet d'étudier le signal provenant majoritairement des sources rostrales (supérieures) orientées verticalement (lemnisque latéral, colliculus inférieur ; voir partie précédente). Le *montage horizontal* (mastoïde référencée à la mastoïde opposée) reçoit quant à lui préférentiellement l'activité des sources caudales (inférieures) radiales (nerf auditif, noyau cochléaire ; Galbraith *et al.*, 2001), constituée de composantes spectrales plus hautes-fréquences du fait de la limite de suivi de fréquence (phase-locking) plus élevée inhérente à ces structures.

Des développements méthodologiques ont récemment montré l'intérêt d'utiliser un système EEG doté de 32 voies d'enregistrement ou plus dans les études de Speech ABR, non seulement pour l'enregistrer simultanément aux réponses corticales (Musacchia *et al.*, 2008), mais également pour bénéficier de techniques d'analyses multi-électrodes. Celles-ci peuvent permettre de diminuer la durée d'enregistrement de la Speech ABR (Bharadwaj & Shinn-Cunningham, 2014), d'accéder à sa topographie (Bidelman, 2015b), ou encore de la débruiter (de Cheveigné & Simon, 2008).

Nombre d'essais enregistrés

L'obtention de la Speech ABR (et des potentiels évoqués en règle générale) passe par le calcul de la moyenne d'un nombre d'essais unitaires (réponse à un seul *stimulus*). La quantité d'essais nécessaires dépend du niveau de bruit et de l'amplitude de l'activité d'intérêt. Aussi, cette quantité s'associe aux paramètres de durée du *stimulus* et de cadence de stimulation pour déterminer le temps d'acquisition.

La Speech ABR étant inférieure d'un ordre de grandeur aux PEA corticaux, il faut théoriquement faire la moyenne d'environ 2000 essais unitaires pour obtenir une réponse avec un rapport signal-sur-bruit (Signal-to-noise ratio, SNR) correct, là où les PEA corticaux n'en nécessiteraient que 200. Ainsi, pour un *stimulus* de 180 ms avec un ISI de 70 ms, soit une cadence de $1000/(180+70) = 4$ *stimuli* par seconde, il faudra $2000/4 = 500$ secondes pour recueillir la Speech ABR (soit un peu plus de 8 minutes).

Comme le niveau de bruit peut varier d'un environnement d'acquisition à l'autre, et la qualité de la Speech ABR d'un individu à l'autre, le nombre d'essais requis pour observer une réponse satisfaisante n'est pas fixe. Ainsi, il est possible de concevoir un système d'arrêt de l'acquisition lorsque le SNR (par exemple) atteint un certain seuil, ce qui peut se produire dès 800 essais dans un environnement électromagnétique propre et chez un sujet calme présentant une réponse robuste.

3.2.1.3 Traitements et analyses

Traitement des données

- Filtrage

Pouvant être effectué pendant ou après l'acquisition (*online* ou *offline* respectivement), le filtrage des données permet d'étudier certaines fréquences isolément. Dans le cas de la Speech ABR, la bande spectrale d'intérêt définie entre 80 et 3000 Hz est isolée du signal par un filtre passe-bande. La fréquence de coupure inférieure varie entre 70 et 100 Hz selon les études, l'idée étant d'isoler la Speech ABR des réponses corticales plus basses en fréquences. La fréquence de coupure supérieure dépend, outre de la fréquence d'échantillonnage du système EEG utilisé, de l'activité de suivi de fréquence maximale observée tout du moins dans les sources radiales visibles en utilisant un montage horizontal. Cependant, la limite de phase-locking du colliculus inférieur est communément admise comme située entre 1000 et 1500 Hz, donc une fréquence de coupure supérieure à 1500 Hz apportera surtout du bruit pour une activité observée en montage vertical.

- Artéfacts

Les données d'EEG peuvent contenir des artéfacts de différents types, le plus souvent liés au rayonnement du courant alternatif du secteur électrique (50 Hz ou 60 Hz, par exemple en France et aux Etats-Unis), et à l'activité musculaire des yeux et des muscles crâniens. L'artéfact du secteur est visible dans le spectre du signal EEG sous la forme d'un peigne fréquentiel composé du 50 Hz et de ses harmoniques (impaires uniquement, pour notre système d'acquisition), et peut être nettoyé en utilisant des filtres de type « notch », permettant de s'attaquer à une fréquence précise. L'activité musculaire liée aux yeux se manifeste de façon prépondérante aux électrodes préfrontales (Fp1 et Fp2), mais n'est plus spécialement visible après filtrage des données par le passe-bande décrit au point précédent. L'activité des muscles crâniens s'observe surtout au niveau des électrodes temporales (T7 et T8) ; elle est principalement haute fréquence, et elle demeure donc dans le signal après filtrage ; un moyen de s'en débarrasser est de rejeter les essais unitaires dépassant un certain seuil (amplitude, ou déviation par rapport à la moyenne).

Deux artéfacts viennent s'ajouter à cette liste dans le cas de la Speech ABR : l'artéfact de stimulation, et l'artéfact du muscle post-auriculaire (PAM pour *Post-Auricular Muscle*). Nous avons précédemment abordé le premier ; il s'observe à une latence quasi nulle, et est la trace du *stimulus* dans les données en conséquence du rayonnement électromagnétique des câbles des écouteurs ; il peut être minimisé en faisant la moyenne des réponses aux polarités opposées. L'artéfact PAM est d'une nature bien plus étonnante : à forte intensité de stimulation, le muscle post-auriculaire est capable de suivre la fréquence fondamentale du

stimulus (Sohmer *et al.*, 1977 ; O'Beirne & Patuzzi, 1999) ! Il est observé exclusivement sur les électrodes mastoïdiennes, primordiales du fait de leur rôle de référence pour l'étude de la Speech ABR. Il n'existe actuellement pas de technique pour s'en débarrasser aisément, si ce n'est un changement de référence pour une référence moyenne après avoir supprimé des données les électrodes mastoïdiennes atteintes. Dans les travaux expérimentaux de cette présente thèse, un participant ayant effectué plusieurs sessions d'enregistrement de Speech ABR a montré l'artéfact PAM à chaque session, suggérant qu'il est reproductible pour des paramètres expérimentaux constants.

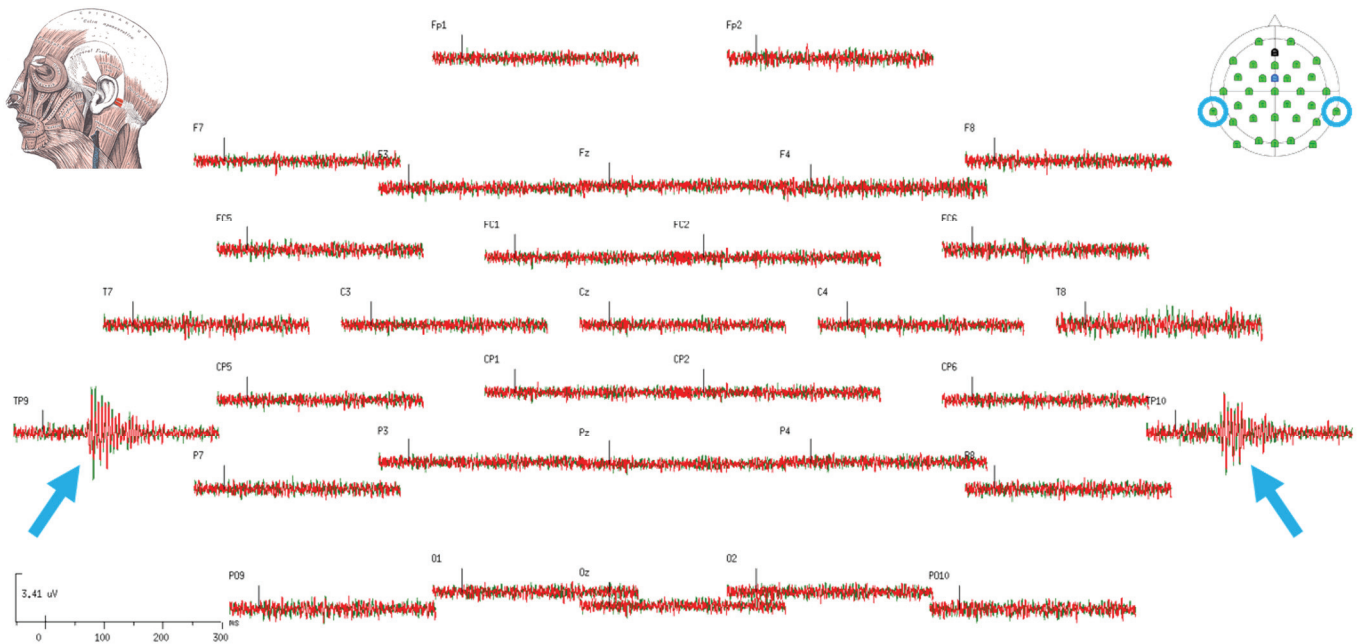


Figure 3.10. Artéfact du muscle post-auriculaire sur les électrodes mastoïdiennes (flèches et cercles bleus), dans les données filtrées et moyennées de la Speech ABR évoquée par une syllabe /ba/. Ces données proviennent d'un patient de l'étude 4, qui a présenté l'artéfact PAM à chacune des 5 sessions d'enregistrement, et pour chacune des polarités (courbes rouges et vertes). Cet artéfact provient du muscle post-auriculaire, fixé à l'apophyse mastoïde de l'os temporal (en rouge sur le schéma en haut à gauche).

- Découpage des essais et moyenne

Une fois filtrées et débarrassées des différents artéfacts, les données sont découpées en essais (*trials*) ou époques (*epochs*), désignant des morceaux de signal découpés autour du début du *stimulus*. Pour la Speech ABR ayant une latence entre 6 et 20 ms, et par exemple pour un *stimulus* de 250 ms, il est convenable de découper le signal entre -50 et 350 ms. Nous comprenons alors l'influence de la cadence de stimulation et de l'ISI, en cela qu'il faut laisser suffisamment d'espace entre deux *stimuli* consécutifs pour que le début du *stimulus* $n+1$ n'empiète pas sur la fin de la Speech ABR au *stimulus* n .

Une correction de ligne de base peut être effectuée, permettant de ramener à zéro microvolt l'amplitude moyenne d'une période qui serait définie entre -50 et 0 ms dans notre exemple, et ainsi d'aligner tous les essais sur une même amplitude de départ.

La moyenne des essais peut ensuite être calculée, en implémentant communément un rejet sur critère de seuil d'amplitude : tout essai ayant dépassant plus ou moins 50 μV sera rejeté, considéré comme pollué par de l'activité musculaire. Une autre stratégie est le rejet sur gabarit : tout essai s'écartant de plus de 2 écarts-types (par exemple) de la moyenne sera considéré comme anormal, et sera alors rejeté.

- Combinaison des polarités

Si l'information de polarité des *stimuli* a été conservée jusque-là, nous disposons à ce stade d'une moyenne par polarité. Pour s'assurer de la présence d'un artéfact de stimulation et pour observer la réponse la plus robuste, ces signaux sont le plus souvent moyennés, donnant ainsi l'*envelope FFR*. Mais il est possible de moyennner la réponse moyenne à une polarité avec l'opposé de la réponse moyenne à l'autre polarité, pour obtenir la *spectral FFR*.

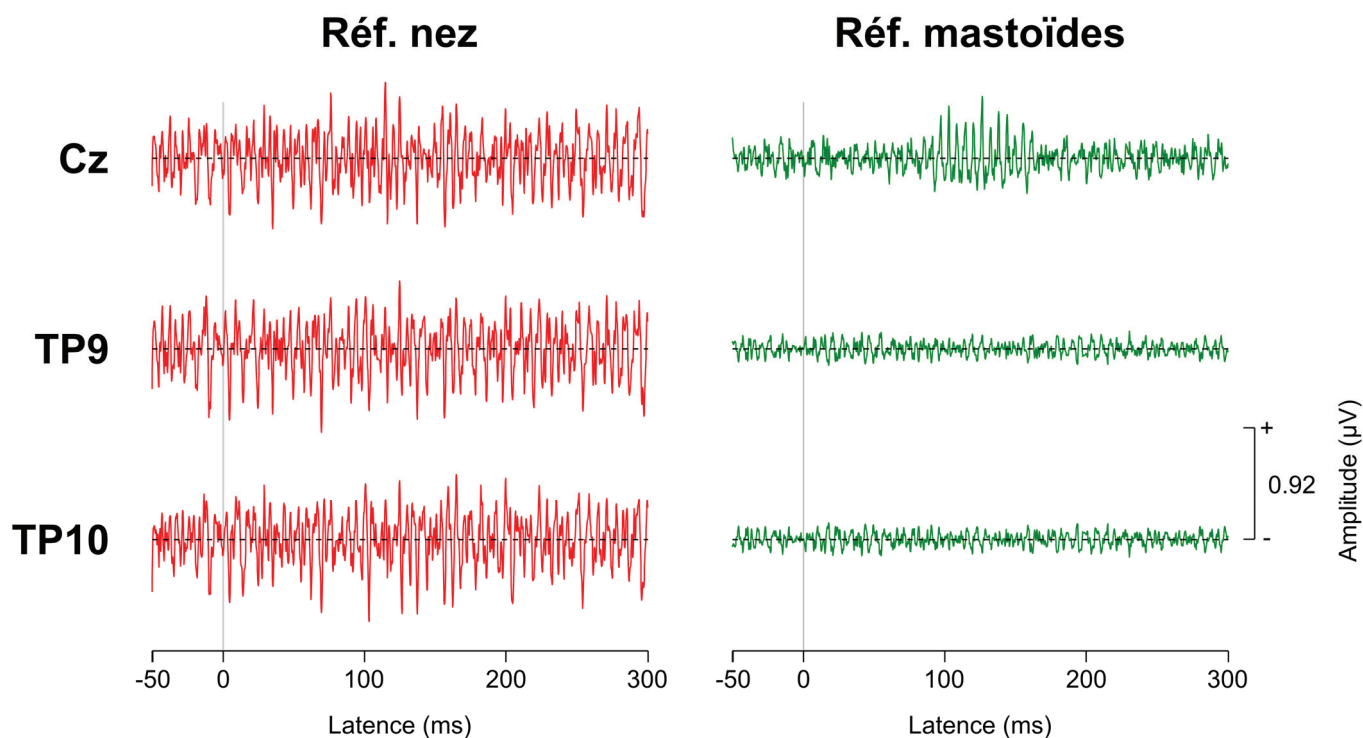


Figure 3.11. Reréférencement des électrodes à la moyenne des mastoïdes, pour les données d'un participant de l'étude 4 en réponse à une syllabe /ba/. L'échelle d'amplitude est commune à tous les tracés. Les données sont précédemment filtrées, découpées en essais, et moyennées avec rejet sur gabarit. Avant reréférencement (courbes rouges à gauche), du bruit est commun à toutes les électrodes, en plus de l'activité qui leur est propre. L'amplitude du bruit est diminuée après reréférencement (courbes vertes à droite), environ de moitié pour Cz.

- Reréférencement

Dans le cas où la référence n'est pas placée aux mastoïdes pendant l'enregistrement, il est délicat d'observer un signal en Cz malgré le bon déroulement des étapes précédentes. Cela vient du fait que la Speech ABR subit une inversion de polarité sur un axe vertical : à Cz s'observe le signal sous forme d'ondes positives (soit y), et aux mastoïdes le même signal sous forme d'ondes négatives (soit $-y$). En reréférencant Cz aux mastoïdes, c'est-à-dire en soustrayant le signal des mastoïdes à celui enregistré en Cz (soit $y - (-y) = 2y$), nous amplifions alors ce signal tout en réduisant le niveau de bruit, ayant la même polarité entre toutes les électrodes.

- Techniques multi-électrodes (ICA, PCA, DSS)

Avant d'explorer les différentes analyses possibles, notons l'intérêt d'enregistrer la Speech ABR sur un nombre conséquent d'électrodes (par exemple 32, au lieu de 3 ou 4) : le signal évoqué par le *stimulus* de parole sera observable non plus seulement sur Cz et les mastoïdes, mais aussi sur les 29 autres électrodes. Des techniques multi-électrodes peuvent alors être employées, telles que le débruitage par séparation des sources (Denoising Source Separation ; de Cheveigné & Simon, 2008) ou l'analyse en composantes principales (PCA), et permettent d'augmenter significativement le SNR (et donc de réduire le nombre d'essais nécessaires à l'obtention d'une Speech ABR de qualité donnée), ainsi qu'éventuellement de séparer certaines activités composant la Speech ABR (telles que celles accessibles par les montages vertical et horizontal). Aussi, les artefacts oculaires peuvent être retirés par l'ICA, ce qui est davantage utile pour les PEA corticaux que pour la Speech ABR, qu'ils impactent moins du fait du filtrage des basses fréquences.

Enfin, l'utilisation de 32 électrodes ouvre l'accès à l'étude des topographies, soit de la répartition du potentiel électrique sur le scalp. Or, cette répartition est fonction des générateurs sous-jacents, et apporte donc une information d'intérêt premier dans l'étude des générateurs neuronaux de la Speech ABR (voir les études 1 et 2 du présent manuscrit).

Analyses

- Latences et amplitudes des pics

Une fois les différentes composantes de la réponse caractérisées par confrontation avec le *stimulus*, une première analyse peut s'effectuer sur la latence et l'amplitude des pics remarquables (*onset*, *offset*, pics de la FFR). Elle permet alors d'établir des différences potentielles entre conditions expérimentales.

- Latence de la FFR

La latence de la FFR dans sa globalité (par opposition à la mesure de la latence de ses pics isolés) est souvent mesurée par corrélation croisée entre le *stimulus* et le signal évoqué, du fait de leur importante familiarité.

- SNR

Le rapport signal-sur-bruit (SNR) est un indice de la qualité de la réponse, influencée à la fois par les conditions d'enregistrement et par la fonction même des générateurs neuronaux du signal mesuré. Dans les travaux expérimentaux de la présente thèse, nous l'avons calculé en utilisant la formule : $20 \cdot \log_{10}((E_{XY} - E_Y)/E_Y)$, avec E_Y représentant l'énergie dans une fenêtre temporelle ne contenant pas de signal (et donc seulement du bruit), et E_{XY} l'énergie dans une fenêtre contenant du signal (définie pour englober l'*onset response* et le début de la FFR). Le SNR est notamment utilisé pour estimer l'amplitude de la FFR dans sa globalité.

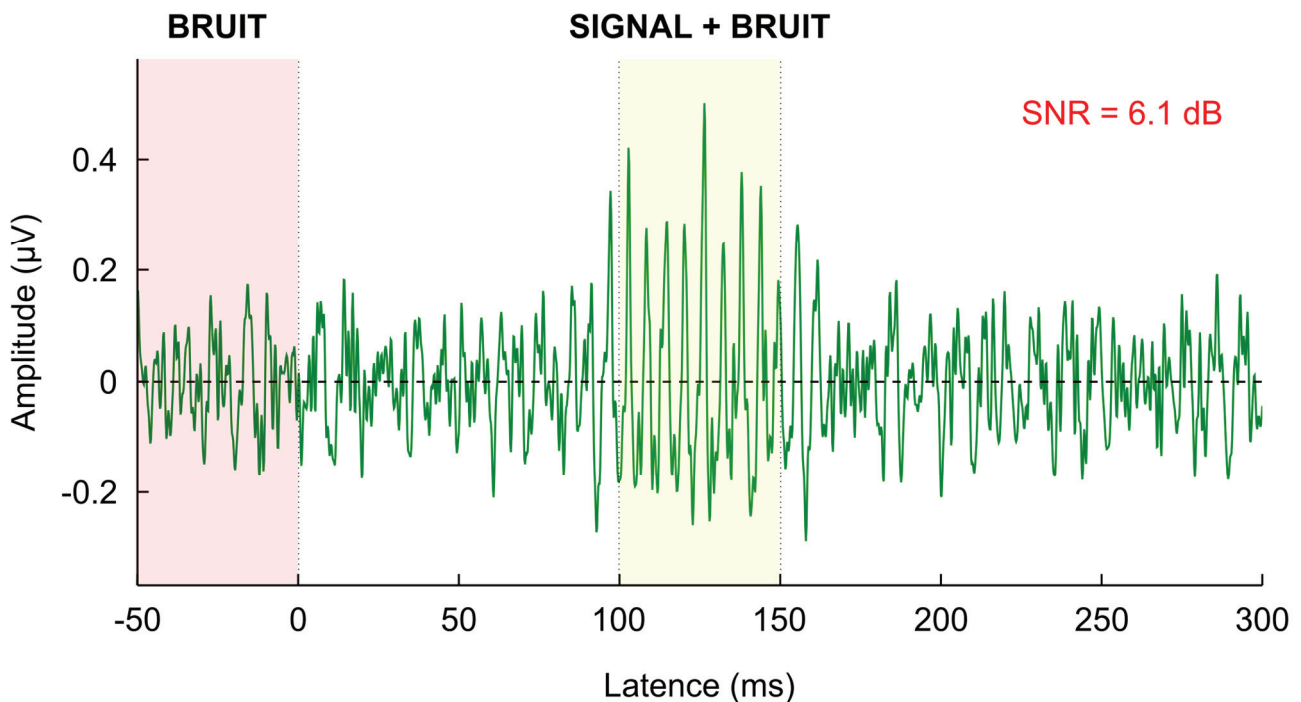


Figure 3.12. Rapport signal-sur-bruit de la Speech ABR présentée Figure 3.11. (en réponse à une syllabe /ba/ chez un participant de l'étude 4).

- Différences entre Speech ABR

Afin de mettre en évidence des différences entre Speech ABR évoquées par différents *stimuli* ou issues de différentes conditions, une analyse se basant sur les différences de phase

en fonction du temps et de la fréquence a été développée, sous le nom de cross-phaseogram (Skoe *et al.*, 2011). Elle évite notamment d'avoir à effectuer un choix arbitraire des pics pour lesquels il serait intéressant de mesurer la latence et l'amplitude, et a pour corollaire l'apport d'une information plus exhaustive quant à la description spectro-temporelle des différences entre Speech ABR.

- Consistance

Deux stratégies permettent d'estimer la reproductibilité de la réponse au fil des essais d'un même enregistrement. Dans la première, deux sous-moyennes sont calculées (essais pairs vs. impairs, ou première moitié vs. seconde moitié), et c'est leur coefficient de corrélation qui mesure la consistance de la réponse (Hornickel *et al.*, 2012b ; Tierney *et al.*, 2013b ; Hornickel & Kraus, 2013). La deuxième stratégie utilise la valeur de verrouillage de phase (phase-locking value), qui donne directement l'information de la reproductibilité du signal (Lachaux *et al.*, 1999 ; Bharadwaj & Shinn-Cunningham, 2014).

- Test-retest

La comparaison des Speech ABR enregistrées au cours de deux sessions espacées de 50 jours environ n'a révélé aucune différence notable, suggérant ainsi une très bonne stabilité de la réponse à travers les sessions (Song *et al.*, 2011a). L'étude en question a reproduit ce constat avec deux groupes de participants, l'un stimulé avec une syllabe /da/ de 170 ms présentée dans le calme ou dans le bruit au cours de deux sessions espacées de 58 jours, l'autre avec le fameux /da/ de 40 ms présenté dans le calme au cours de deux sessions espacées de 41 jours.

Une autre étude a cette fois-ci étudié ce marqueur chez un groupe d'enfants âgés de 8 à 13 ans, sur une période d'un an, encore une fois sans trouver de changement dans les différentes composantes et dimensions de la Speech ABR (Hornickel *et al.*, 2012b).

Ces deux études ouvrent ainsi une fenêtre temporelle de 12 mois minimum pour l'investigation d'effets en paradigme longitudinal.

- Algorithmes

Des algorithmes de détection automatique des réponses ont été développés pour automatiser l'étiquetage de certains pics (Jeng *et al.*, 2011c ; Bidelman, 2014), en s'affranchissant de la subjectivité de l'expérimentateur. L'utilisation diagnostique de la Speech ABR est alors grandement simplifiée, et peut ainsi être adoptée plus aisément en routine clinique, par exemple dans le dépistage de la surdité chez le nourrisson (Jeng *et al.*, 2013).

D'autres algorithmes permettent la classification automatique des voyelles se basant sur les amplitudes spectrales, et atteignant une précision de 78,3 % (Sadeghian *et al.*, 2011), pouvant alors servir d'indice objectif des confusions phonémiques à l'échelle de l'encodage neuronal ; ou encore la modélisation de la FFR, en extrayant automatiquement cinq paramètres pour la décrire, de façon robuste y compris avec un faible SNR (Jeng *et al.*, 2011a).

En automatisant l'analyse de données, ces techniques sophistiquées facilitent l'utilisation généralisée de la Speech ABR qui, comme nous allons le voir, a un fort potentiel clinique dans de nombreuses thématiques liées plus ou moins directement à l'audition et au langage.

- Statistiques

Les mesures de latence, d'amplitude et de SNR peuvent entrer dans des analyses de variance classiques. Des analyses de corrélations peuvent mettre en évidence des disparités entre Speech ABR à différents *stimuli* ou de différentes conditions expérimentales (par exemple dans les études de test-retest ou pour la mesure de la consistance du signal). Aussi, des tests de Wilcoxon peuvent comparer la variance de chaque échantillon du signal de la Speech ABR à celle d'une ligne de base (souvent la période pré-*stimulus*), pour évaluer statistiquement l'émergence de la réponse par rapport au bruit.

3.2.2. Générateurs neuronaux

3.2.2.1. Etat des connaissances

Jusqu'à présent, aucune étude n'a apporté de preuve forte renseignant sur l'identité exacte des générateurs neuronaux de la Speech ABR. Aussi, bien rares sont les articles ayant abordé la question, puisque nous n'en avons dénombré que trois. Cela peut s'expliquer par l'existence d'un dogme selon lequel la Speech ABR serait générée au niveau sous-cortical, étant donnée la ressemblance entre son *onset response* et l'onde V des Click ABR d'une part, et sa FFR et « la » FFR d'autre part (ces deux réponses classiques de l'électrophysiologie de l'audition ayant effectivement fait l'objet de très nombreuses études pointant vers des générateurs neuronaux périphériques et sous-corticaux).

Parmi les trois articles s'étant penchés sur la question des générateurs neuronaux de la Speech ABR, le premier a comparé les réponses évoquées par des sons de complexité progressive (du ton pur à la syllabe), et s'est basé sur les propriétés d'encodage des noyaux et cellules du tronc cérébral pour arriver à la conclusion d'une FFR générée au niveau du colliculus inférieur (Akhoun *et al.*, 2010). Le second a retracé l'histoire des générateurs de la FFR aux tons chez l'animal, et du comportement de la FFR par rapport à l'influence de

certaines effets et en comparaison aux comportements des réponses corticales et cochléaires, pour conclure à des générateurs sous-corticaux de la Speech ABR, nommément le noyau cochléaire, le lemme latéral et le colliculus inférieur (Chandrasekaran & Kraus, 2010b). Le plus récent s'est intéressé aux topographies de la Speech ABR, en s'en servant pour alimenter une reconstruction de source *via* deux méthodes (dipôles et technique de la trajectoire à trois voies de Lissajous) : les sources ont été localisées au niveau du tronc cérébral supérieur, dans le colliculus inférieur (Bidelman, 2015b).

Ainsi, à ce jour, les rares résultats récents ont confirmé une génération sous-corticale de la Speech ABR, appuyés par un raisonnement basé sur la littérature des Click ABR et de la FFR.

Avant d'entrer dans le récit de cette littérature, il est important de bien rappeler les différentes activités regroupées sous l'étiquette *Speech ABR*. Constituant un premier type de composante, les réponses transitoires, dont l'*onset* et l'*offset response*, permettent l'encodage des variations rapides d'énergie des sons de parole ; elles s'observent avec une latence de 6 à 10 ms. Le second type de composante est la FFR, réponse soutenue permettant l'encodage du contenu spectral des *stimuli* (F0, harmoniques, formants) ; dans le cas présent, soit en réponse à des sons de parole, elle s'observe avec une latence plus tardive, de 9 à 15 ms, relativement à la FFR aux tons (ce qui, en outre, n'exclut pas la présence d'autres pics arrivant plus tardivement).

Comme nous l'avons vu dans la partie précédente, certains paramètres d'acquisition et de traitement des données influent sur le contenu spectral de la FFR :

- un montage horizontal (mastoïde référencée à la mastoïde controlatérale) favorisera l'observation de sources radiales (nerf auditif et noyau cochléaire) et de leur activité relativement très-haute-fréquence, par rapport à l'activité observée avec un montage vertical (vertex référencé aux mastoïdes, aux lobes d'oreille ou au niveau de la nuque postérieure médiane), imputée au tronc cérébral supérieur (lemme latéral et colliculus inférieur) et permettant d'observer une activité relativement moins haute-fréquence ;
- la FFR spectrale (soustraction des polarités) mettra l'emphasis sur le haut du spectre (autour de 800 Hz, valeur du premier formant F1 de la voyelle /a/, à titre d'exemple), et la FFR d'enveloppe (addition des polarités) sur le bas du spectre (F0 et premières harmoniques).

En résumé, la Speech ABR est un mélange d'activités transitoires et soutenues plus ou moins hautes-fréquences en fonction des paramètres de recueil précités, ce qui suggère en soi la participation de plusieurs générateurs. Comme nous allons le voir dans la partie suivante sur le déroulement des champs d'application de la Speech ABR, une majorité d'études ne

s'intéresse qu'à la FFR d'enveloppe en montage vertical, ce qui s'explique par la robustesse de la réponse (prépondérance de F0 donc facilement observable dans le domaine temporel, et minimisation de l'artéfact de stimulation) ; étant donné le lien accepté entre FFR à la parole en montage vertical et tronc cérébral supérieur, ces études interprètent alors toute variation dans l'encodage de F0 comme une preuve de modulation de l'activité sous-corticale, si ce n'est de plasticité du tronc cérébral auditif supérieur.

3.2.2.2. Click ABR

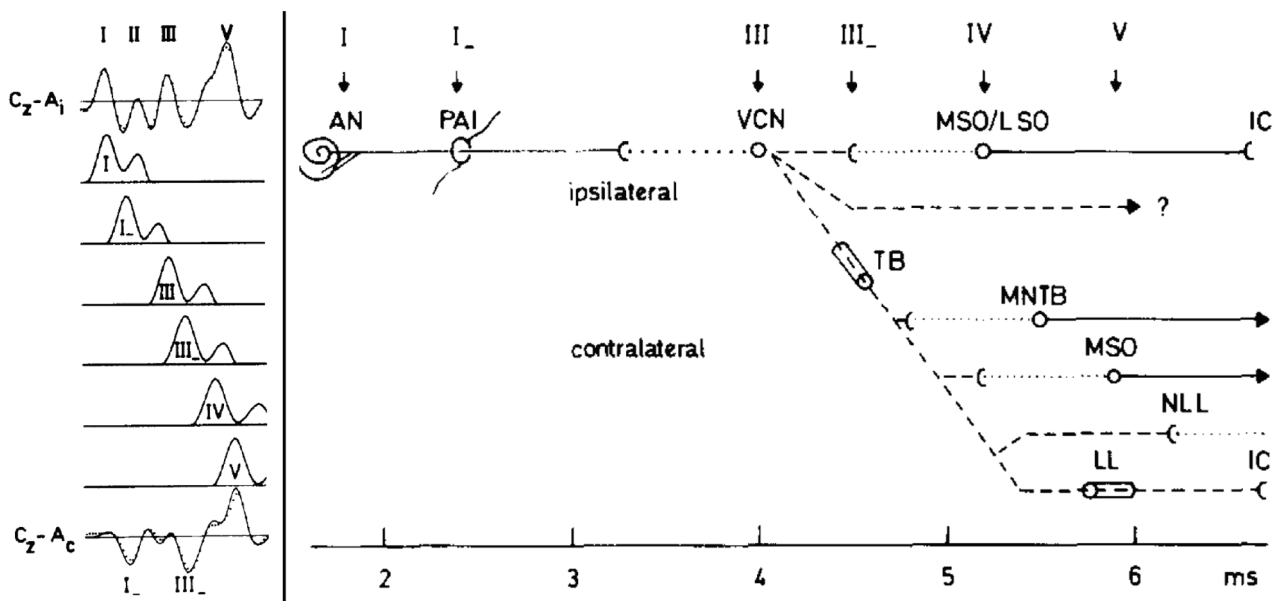


Figure 3.13. Générateurs de la Click ABR (figure adaptée de Scherg & von Cramon, 1985). À gauche, la réponse est présentée pour un montage ipsilatéral à l'oreille stimulée (en haut), pour une référence contralatérale (en bas ; Cz représente le vertex ; Ai et Ac représentent les lobes d'oreille ipsi- et contralatéral, A signifiant *auricular*), et pour chacun des 6 dipôles d'un modèle de reconstruction de sources (au milieu). À droite, les générateurs supposés sont représentés en fonction de la latence (AN : nerf auditif ; PAI : pore acoustique interne, par où passe le nerf auditif pour rejoindre le noyau cochléaire ; VCN : noyau cochléaire ventral ; TB : corps trapézoïde, ou strie acoustique ventrale, permettant de croiser la ligne médiane ; MSO/LSO : noyau médian/latéral du complexe olivaire supérieur ; MNTB : noyau médian du corps trapézoïde, composante du complexe olivaire supérieur ; (N)LL : (noyaux du) lemnisque latéral ; IC : colliculus inférieur).

Après leur découverte autour de l'année 1970 (Jewett *et al.*, 1970 ; Jewett & Williston, 1971), les PEAP aux clics (ou Click ABR) ont été l'objet d'un nombre conséquent d'études visant à en déterminer les générateurs. En premier lieu, différents modèles animaux (chat, singe, cochon d'inde, entre autres) ont permis d'une part la comparaison entre activité enregistrée *in situ* par l'implantation d'électrodes intracrâniennes et activité distale telle qu'enregistrée depuis le scalp (Achor & Starr, 1980a ; Møller & Burgess, 1986), et d'autre part

l'étude des effets de lésions ou d'inactivations ciblées de structures nerveuses périphériques et sous-corticales sur la qualité de la réponse aux clics (Buchwald & Huang, 1975 ; Achor & Starr, 1980b ; Wada & Starr, 1983a,b,c). Chez l'Homme, l'étude de patients présentant des lésions bien localisées (Starr & Achor, 1975 ; Stockard & Rossiter, 1977), les enregistrements *in situ* peropératoires (Hashimoto *et al.*, 1981 ; Møller & Jannetta, 1981, 1982 ; Møller *et al.*, 1981a,b) , et l'utilisation de techniques de reconstruction de source à partir des données de scalp (Scherg & von Cramon, 1985 ; Pratt *et al.*, 1985, 1987), ont mené à une connaissance solide des générateurs de la Click ABR.

Ainsi, les ondes I et II sont générées par le nerf auditif, respectivement au niveau distal ou proximal (par rapport au centre), les ondes III et IV respectivement par le noyau cochléaire ventral (et ses cellules globulaires de type « bushy » capables d'une grande précision temporelle) et le complexe olivaire supérieur, et l'onde V par le lemnisque latéral majoritairement contralatéral (Figure 3.13.).

3.2.2.3. FFR

Découverte à la fin des années 1960 sous forme d'une activité neuronale de même fréquence et forme d'onde que le ton utilisé en *stimulus* (Worden & Marsh, 1968), la FFR a fait l'objet d'investigations semblables à celles décrites pour les Click ABR, soient d'études chez l'animal en comparant la réponse intracrânienne à celle enregistrée sur le scalp, ou en étudiant l'effet de lésions focales (Marsh *et al.*, 1970 ; Smith *et al.*, 1975 ; Gardi *et al.*, 1979 ; Yamada *et al.*, 1980), et d'études chez l'Homme s'intéressant aux effets de déficits sensoriels, de lésions dans le tronc cérébral et de pathologies neurologiques (Sohmer & Pratt, 1977 ; Sohmer *et al.*, 1977 ; Yamada *et al.*, 1982).

Il a ainsi été mis en évidence une participation des potentiels microphoniques cochléaires à la génération de la FFR, avec une latence comprise entre 1 et 2 ms, d'une source nerveuse d'une latence de 4 ms identifiée au noyau cochléaire (du fait du rapprochement à l'onde III des Click ABR ; Figure 3.13.), et d'une autre source nerveuse d'une latence de 6 ms identifiée au tronc cérébral supérieur (colliculus inférieur ou lemnisque latéral). Notons que ces résultats ont été majoritairement obtenus pour des tons de fréquence fondamentale égale à 500 Hz, et devraient donc n'être transférés qu'avec grande prudence aux sons de parole de moindre fréquence fondamentale.

3.2.2.4. Cortex auditif ?

De prime abord, nous pourrions trouver déroutante l'idée d'un point sur l'activité corticale dans cette partie du manuscrit consacrée à la Speech ABR. Pourtant, et en résonance avec la remarque concluant le paragraphe précédent, une étude datant de 1980 a révélé

l'existence chez le singe de PEA corticaux montrant une activité de suivi de phase pour des *stimuli* périodiques (dont des sons de parole) de F0 égale à 100 et 250 Hz ; aucune réponse de ce type n'était cependant retrouvée pour une F0 de 500 Hz (Steinschneider *et al.*, 1980). La ligne de recherche développée par Mitchell Steinschneider a ensuite permis de mieux comprendre la représentation corticale de traits phonétiques tels que le lieu d'articulation (Steinschneider *et al.*, 1995a) ou le voisement (Steinschneider *et al.*, 1995b), toujours chez le primate non-humain.

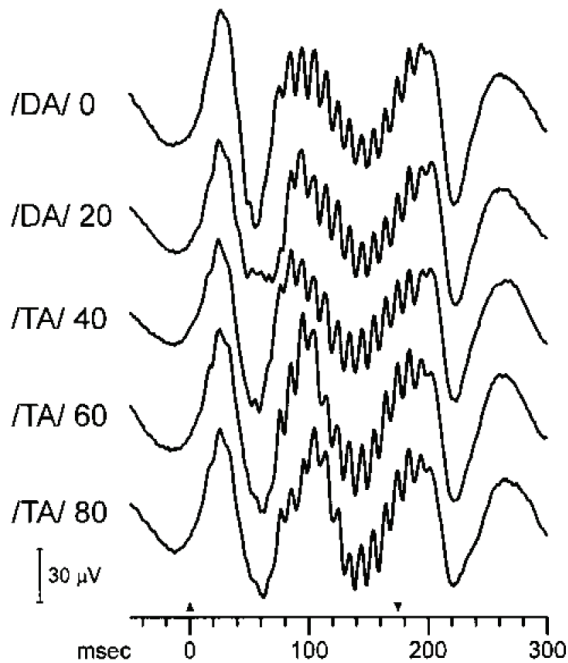
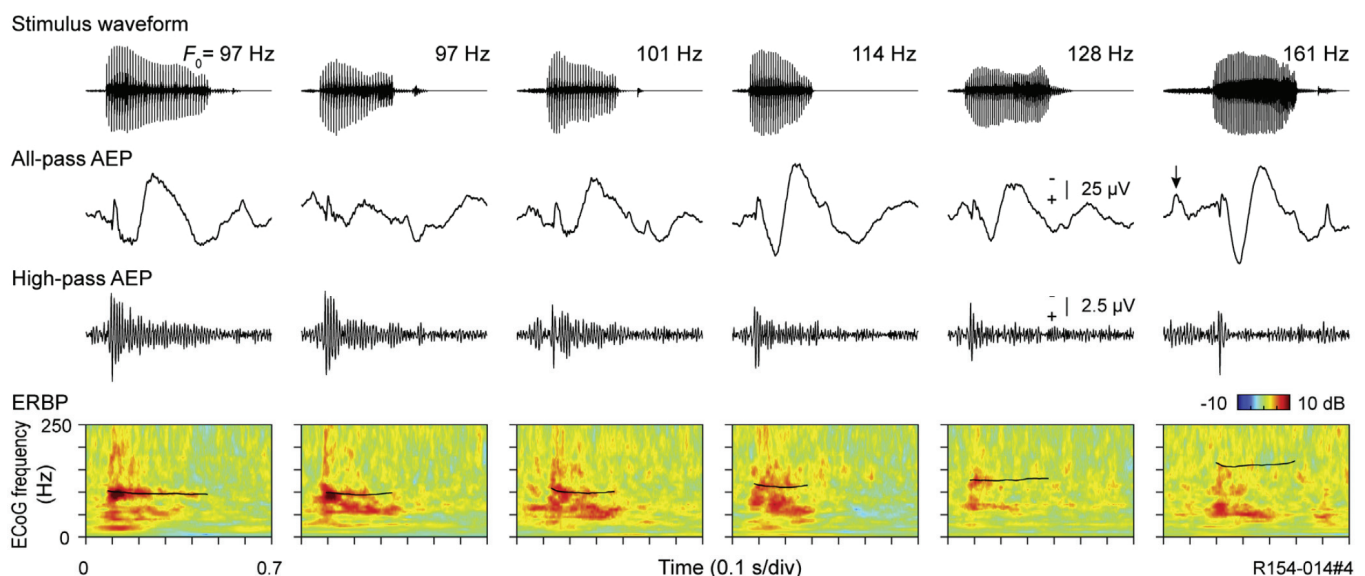


Figure 3.14. FFR évoquée par plusieurs syllabes de F0 égale à 100 Hz (Steinschneider *et al.*, 1999). Ces données ont été enregistrées depuis le gyrus de Heschl d'un patient épileptique implanté.

En 1999, l'encodage temporel du voisement a pu être étudié par l'enregistrement intracrânien de potentiels de champs locaux (*Local Field Potentials*) chez le patient humain épileptique implanté dans le cadre de sa prise en charge clinique. Parmi les résultats de l'étude, une activité de suivi de phase était observée en réponse à des syllabes synthétiques de F0 égale à 100 Hz (Steinschneider *et al.*, 1999), montrant alors l'existence d'une activité FFR dans le cortex auditif, au moins jusqu'à cette fréquence (Figure 3.14.).

Par la suite, cette fréquence limite n'a que très peu été reconsidérée, puisqu'en 2011 une étude du

Figure 3.15. F0 limite de 114 Hz pour l'activité corticale de suivi de phase (Nourski & Brugge, 2011).



même groupe a montré une FFR corticale à 114 Hz, mais absente pour des F0 plus élevées, en l'occurrence à 128 et 161 Hz (Figure 3.15. ; Nourski & Brugge, 2011 ; voir Steinschneider *et al.*, 2013 pour une revue).

Pour synthétiser les différents résultats de cette partie et les remettre en relation avec la Speech ABR, il est très clair que les voies auditives sous-corticales présentent des activités transitoires et soutenues. Cependant, pour la FFR évoquée par F0, c'est-à-dire dans notre cas mesurable par un montage vertical et en polarités alternées, une participation corticale dans la génération de cette réponse enregistrée au niveau du scalp n'est pas à exclure, au moins pour des F0 allant jusqu'à 114 Hz. Par opposition, et du fait de cette fréquence maximale, le cortex auditif primaire ne semble pas être capable de contribuer à l'encodage des harmoniques et des formants, tel qu'observé dans la Speech ABR.

3.2.3. Champs d'application

Après avoir présenté la nature de la Speech ABR, ses différentes composantes et générateurs potentiels, et sa méthodologie, nous allons retracer son histoire et son utilisation dans des thématiques scientifiques d'horizons divers.

3.2.3.1. Les débuts d'un marqueur prometteur

Le premier enregistrement connu d'une Speech ABR remonte à 1995, quand Gary Galbraith découvre avec son équipe que le tronc cérébral auditif est capable d'encoder un son de parole avec une précision impressionnante. En effet, en diffusant le signal électrophysiologique *via* des haut-parleurs comme s'il s'agissait d'un signal acoustique, il était alors possible de percevoir le *stimulus* utilisé (Galbraith *et al.*, 1995). Durant les années qui suivent, la même équipe met en évidence un lien entre perception et Speech ABR : l'effet de transformation verbale influe sur l'amplitude de la FFR lorsque le *stimulus* est perçu comme une autre syllabe à force d'être répété (Galbraith *et al.*, 1997) ; puis un effet de l'attention : dans une tâche d'écoute dichotique, la F0 est plus ample dans la FFR du *stimulus* sur lequel l'attention est portée (Galbraith *et al.*, 1998).

En 1999, Ananthanaryan Krishnan démarre une ligne de recherche sur la FFR en réponse à des sons mimant les phonèmes, par un article montrant la présence dans la FFR d'informations sur les formants et sur des fréquences issues de distorsions cochléaires (Krishnan, 1999), suivi l'année suivante d'un deuxième article montrant la capacité de la FFR

à suivre des sons variant en fréquence dans le temps (Krishnan, 2000). Il sera rejoint en 2004 par Jackson Gandour, puis en 2009 par Gavin Bidelman, produisant ensemble et chacun une littérature abondante.

C'est en 2001 que l'équipe de Nina Kraus commence à s'intéresser à la Speech ABR. Ayant déjà travaillé avec les PEA corticaux pour caractériser des effets du vieillissement (ondes P1, N1 et MMN ; Bellis *et al.*, 2000 J Neurosci) ou des troubles des apprentissages (Cunningham *et al.*, 2000 Ear Hear), elle va mettre en évidence plusieurs déficits dans la Speech ABR d'enfants justement atteints de ces troubles : transition du premier formant dégradée, fidélité au *stimulus* diminuée, et *onset response* retardée et réduite en amplitude, notamment lorsque le *stimulus* est présenté dans le bruit (ce *stimulus* étant la syllabe /da/ synthétique de 40 ms présentée ci-dessus).

A partir de ce moment, les études utilisant la Speech ABR se multiplient petit à petit, menées par les trois groupes de recherche précités suivant chacun leur ligne (liens avec la cognition pour Galbraith, linguistique et phonétique pour Krishnan, et caractérisation des troubles des apprentissages et de l'audition de la parole dans le bruit pour Kraus) jusqu'à l'explosion depuis l'année 2005 de la quantité d'articles et de thématiques abordées, avec par ordre d'apparition : entraînement auditif et plasticité neuronale, représentation des différentes caractéristiques des sons, expertise linguistique et bilinguisme, interactions audiovisuelles, expertise musicale, autisme, émotions, attention, vieillissement, développement de la fonction auditive, liens avec d'autres types de mesures, études factorielles, et plus récemment troubles de l'audition, appareils auditifs, épilepsie, cécité, amusie congénitale, schizophrénie et hyperactivité (thématiques auxquelles viennent s'ajouter celles précitées).

Nous allons à présent explorer les résultats principaux ayant rythmé ces différentes thématiques pour ainsi obtenir une vue d'ensemble du domaine d'étude. Nous terminerons alors la partie introductive de cette thèse en dressant un état des lieux synthétique des connaissances sur la Speech ABR résultant de vingt années de recherche.

3.2.3.2. Audition

Asymétrie du système auditif

Le traitement de la parole et plus généralement du langage dans le cerveau humain est latéralisé, avec une dominance de l'hémisphère gauche dans le traitement temporel fin des informations formantiques, et du droit pour les variations plus lentes liées par exemple au rythme des syllabes (Zatorre & Belin, 2001).

Des liens entre Speech ABR et latéralisation corticale du traitement de la parole ont été mis en évidence : la précision temporelle de la Speech ABR est corrélée au traitement cortical

latéralisé des informations acoustiques rapides (Abrams *et al.*, 2006), alors qu'elle n'a pas de lien avec le traitement des informations lentes telles que le rythme syllabique (Abrams *et al.*, 2010). Aussi, pour une syllabe tonale donnée et la même syllabe avec trajectoire de la F0 inversée (donc linguistiquement non-pertinente), la FFR au premier *stimulus* est plus grande lorsque celui-ci est présenté à l'oreille droite (et donc traité dans l'hémisphère gauche), suggérant une asymétrie des voies auditives ascendantes en amont du cortex auditif (Krishnan *et al.*, 2011a), en acceptant bien sûr l'idée de générateurs sous-corticaux de la Speech ABR.

Parole et phonétique

De nombreux *stimuli* approximant les sons de parole naturelle ont été étudiés dans leur capacité à évoquer la FFR. Ils permettent également d'étudier de façon analytique les différents indices phonétiques encodés par la Speech ABR : formants F1 et F2 (Krishnan, 1999, 2002 ; Johnson *et al.*, 2008a ; Krishnan & Agrawal, 2010), trajectoires fréquentielles (Krishnan & Parkinson, 2000 ; Krishnan *et al.*, 2004), et fréquence fondamentale liée au pitch (Krishnan & Plack, 2011 ; Swaminathan *et al.*, 2008).

La Speech ABR encode d'ailleurs les différents indices phonétiques si bien qu'en la sonorisant, c'est-à-dire en transformant le signal électrophysiologie en son, il est possible de percevoir le *stimulus* l'ayant évoquée (Galbraith *et al.*, 1995 ; Weiss & Bidelman, 2015).

Parole dans le bruit

De nombreuses études pour la plupart menées par l'équipe de Nina Kraus ont progressivement institué la Speech ABR comme un prédicteur performant des capacités de compréhension de la parole dans le bruit (Anderson *et al.*, 2013b). Pour une personne comprenant mal la parole dans le bruit, l'idée générale est la suivante : la qualité de l'encodage neuronal d'un *stimulus* de parole est dégradée par le bruit, entraînant une représentation neuronale imprécise de ce *stimulus*, et donc une baisse de son audibilité (Anderson & Kraus, 2010a,b ; Song *et al.*, 2011b).

Chez des enfants avec troubles de compréhension dans le bruit, la Speech ABR évoquée par une syllabe dans le calme ne présente pas de déficit temporel par rapport à celle d'enfants normo-entendants, mais une syllabe diffusée dans le bruit engendre une réponse dégradée et retardée, spécifiquement au niveau de la transition formantique, plus sensible au bruit des points de vue acoustique et comportemental (Anderson *et al.*, 2010a). Un effet du bruit dépendant de la fréquence a aussi été mis en lumière, avec la même logique : l'encodage était le plus dégradé au niveau des bandes de fréquences correspondant aux parties encodées du *stimulus* (basses pendant la transition formantique, hautes pendant la voyelle ; Tierney *et al.*, 2011).

Concernant l'encodage spectral, un autre groupe d'enfants avec troubles de compréhension dans le bruit a été testé, les Speech ABR desquels présentent un déficit dans l'encodage de F0 et de 3F0 (ou H3, sa deuxième harmonique). L'interprétation proposée était que cette altération entraîne une dégradation de la perception du pitch, utilisé pour cibler la voix d'un locuteur dans un environnement bruyant (Anderson *et al.*, 2010b). Aussi, le degré de modulation de l'encodage spectral par le contexte de répétition (réponse au *stimulus* présenté seul, contre réponse au *stimulus* accompagné de 7 autres *stimuli* en quantités égales) a été établi comme un prédicteur des capacités de compréhension dans le bruit (Chandrasekaran *et al.*, 2009).

Dans une étude chez l'adulte âgé comprenant mal dans le bruit (participants entre 60 et 73 ans), l'encodage de F0 est aussi moins robuste, et la réponse à un *stimulus* présenté dans le bruit est fortement dégradée par rapport à la réponse au même *stimulus* présenté dans le calme (Anderson *et al.*, 2011).

Des relations ont également été établies entre Speech ABR et PEA corticaux dans le bruit, avec pour la première réponse les effets classiquement observés de retard de latence, diminution d'amplitude et baisse de la précision neuronale, et pour les seconds des ondes N1 augmentée et P2 diminuée. Ce résultat met ainsi en évidence une interaction entre ces deux niveaux de traitement des sons de parole, respectivement sous-cortical et cortical (en considérant effectivement que la Speech ABR soit sous-corticale ; Parbery-Clark *et al.*, 2011a).

Enfin, le SNR d'un *stimulus* de parole en-dessous duquel la F0 de la Speech ABR correspondante deviendra significativement dégradée a été située à 0 dB. Cela fixe une limite à ne pas franchir lors des expérimentations utilisant la Speech ABR dans le bruit, afin de s'assurer de l'obtention d'une réponse détectable (Li & Jeng, 2011).

Entraînement auditif et plasticité neuronale

Les premiers effets d'un entraînement auditif sur l'encodage neuronal des caractéristiques spectro-temporelles des sons de paroles par la Speech ABR ont été mis en évidence en 2005, avec une plus grande robustesse au bruit de la FFR après un entraînement de 2 mois, consistant en des tests phonologiques et auditifs sous forme de jeux d'ordinateur (Russo *et al.*, 2005). Cette étude succédait alors à une précédente qui n'avait observé ces effets que sur les PEA corticaux (Hayes *et al.*, 2003b).

Plusieurs études ont montré par la suite des effets d'un entraînement sur le suivi de F0 dans la FFR, particulièrement pour les sons de langue tonale les plus complexes (Song *et al.*, 2008b ; Kraus & Chandrasekaran, 2010) et les transitions formantiques (Song *et al.*, 2012 ; Kraus, 2012) ; et sur la robustesse de la Speech ABR dans le bruit (Filippini *et al.*, 2012 ; Anderson *et al.*, 2014).

En complément de ces effets observés pour un entraînement de plusieurs semaines, une plasticité neuronale à plus court terme a également été montrée, cependant pour des tons complexes et non des sons de parole. Dans une première étude sur l'effet du contexte, une mélodie de cinq notes était délivrée aux sujets de façon répétée durant une heure et demie, laquelle mélodie comprenait deux notes répétées, permettant alors d'étudier l'effet d'une répétition globale vs. locale. La répétition globale entraînait une augmentation en amplitude de la première harmonique 2F0 dans la FFR, d'autant plus importante pour la note répétée localement (Skoe & Kraus, 2010b). Une seconde étude sur l'effet du contexte a montré un effet rapide d'apprentissage implicite, avec un encodage des basses fréquences diminué en amplitude pour une séquence de tons complexes structurée plutôt qu'aléatoire (Skoe *et al.*, 2013). L'interprétation commune à ces études est celle d'une plasticité du tronc cérébral auditif induite par l'entraînement ou le contexte de stimulation, interprétation en outre étayée par une comparaison aux mécanismes de plasticité sous-corticale chez l'animal (Tzounopoulos & Kraus, 2009). Nous verrons dans les sections correspondantes les effets des expertises musicale et linguistique, assimilables à un entraînement auditif chronique.

Développement

Dans une première approche, les Speech ABR d'enfants âgés de 3 à 4 ans ont été comparées à celles d'enfants âgés de 5 à 12 ans : les plus âgés présentaient alors des latences de pics plus précoces, et une plus grande amplitude des composantes spectrales aux fréquences contenues dans le *stimulus* ; il n'y avait cependant pas de différence au niveau des Click ABR, suggérant alors une maturation spécifique aux sons complexes au cours de l'enfance (Johnson *et al.*, 2008b). Cette trajectoire développementale commencent dès les premiers mois de vie (Jeng *et al.*, 2010 ; Anderson *et al.*, 2015 ; White-Schwoch *et al.*, 2015) et peut encore s'observer à l'adolescence avec une réduction en amplitude de composantes spectrales marquant le passage à une fonction auditive mature (Krizman *et al.*, 2015).

Vieillesse

Le vieillissement s'accompagne de déficits multiples, et la fonction auditive est bien loin d'être épargnée. Une première étude de Speech ABR a mis en évidence des composantes retardées et diminuées en amplitude, particulièrement aux latences correspondant aux zones de transitions rapides dans le *stimulus* de parole, ainsi que des réponses moins consistantes à travers les essais unitaires, se traduisant par une diminution de l'indice de verrouillage de phase (Figure 3.16. ; Anderson *et al.*, 2012). Des études suivantes ont reproduit ces résultats, avec en règle générale des amplitudes réduites et des latences allongées (Clinard & Tremblay, 2013 ; Clinard & Cotter, 2015).

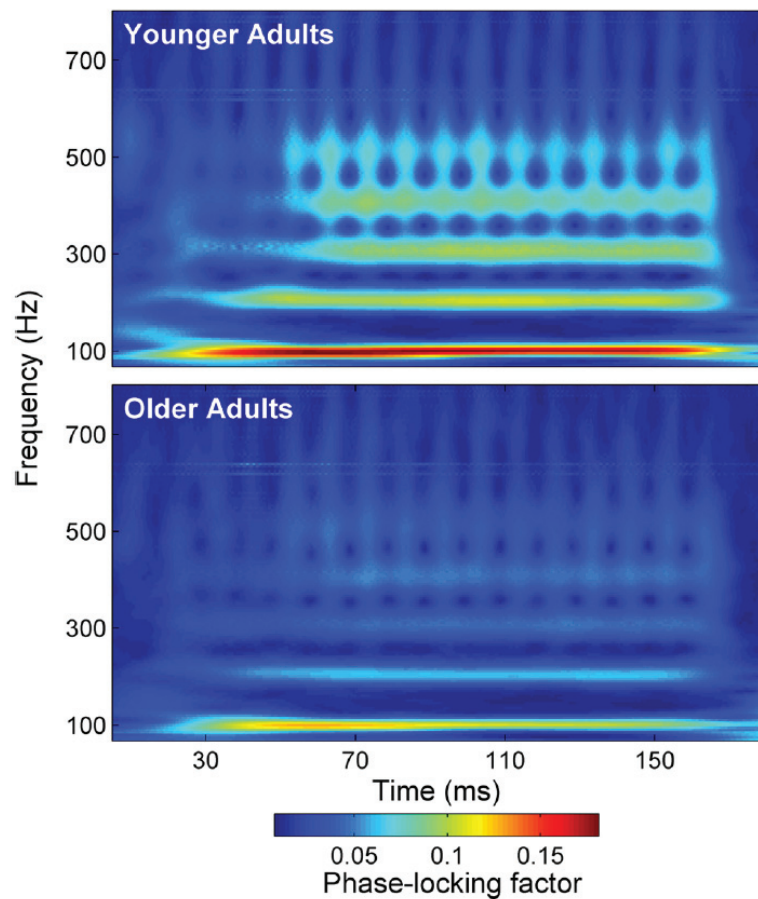


Figure 3.16. Dégradation de la reproductibilité des essais unitaires chez la personne âgée. Cet effet s’observe sur l’ensemble du spectre, suggérant ainsi une imprécision temporelle générale (Anderson *et al.*, 2012). L’indice de verrouillage de phase (*phase-locking factor*) représente le degré de consistance de la réponse, ou plus concrètement de reproductibilité des essais unitaires entre eux. Il varie entre 0 et 1 selon que ces essais soient tous déphasés ou tous parfaitement superposables, respectivement.

Une diminution des capacités de discrimination fréquentielle (Clinard *et al.*, 2010), ainsi que des troubles de compréhension de la parole dans le bruit (Anderson *et al.*, 2013c ; Shinn-Cunningham *et al.*, 2013 ; Fujihira & Shiraishi, 2015) et de la perception catégorielle (Bidelman *et al.*, 2014a) ont également pu être liés à une qualité dégradée de la Speech ABR chez les individus âgés.

Ces effets délétères de l’âge sur l’encodage neuronal des sons de parole peuvent néanmoins et bienheureusement être réduits par un entraînement auditif, menant à un raccourcissement des latences et à une resynchronisation neuronale (Anderson *et al.*, 2013d). Les corrélats comportementaux sont alors une meilleure intelligibilité de la parole dans le bruit, une mémoire auditive à court terme plus efficace, et une durée de traitement auditif plus courte (temps de réponse réduits). L’entraînement utilisé dans cette étude durait deux mois, à raison d’une heure par jour, cinq jours par semaine (40 heures au total), et s’effectuait au domicile du participant sous forme d’un logiciel proposant six exercices progressifs de discrimination, répétition et mémorisation de syllabes et de mots (*Brain Fitness*, Posit Science).

3.2.3.3. Expertises auditives

Langage

Une ligne de recherche très fournie, animée principalement par l'équipe de Krishnan, a investigué l'influence de l'expérience linguistique sur l'encodage d'indices phonologiques particuliers. Notamment, dans les langues tonales telles que le mandarin, le percept de hauteur de ton (*pitch*) confère un sens aux phonèmes : une syllabe prononcée avec une trajectoire fréquentielle montante aura alors une signification différente de cette même syllabe prononcée avec une trajectoire descendante. Dans le cas des langues non-tonales, la hauteur de ton est liée à la prosodie, à l'état émotionnel et à l'identité du locuteur, mais n'influe pas directement sur le sens du mot perçu.

Un autre indice phonologique permettant de discriminer les langues est l'accent tonique (*stress*), soit une variation d'intensité sonore à certains endroits des mots ou des phrases ; par exemple, le français est bien moins accentué que l'anglais, si bien que les français n'arriveront que difficilement à percevoir cet indice dans la parole d'un locuteur anglais.

L'influence de l'accent tonique sur la Speech ABR n'est pas connue, mais comme il s'agit d'une variation d'intensité sonore, nous pourrions nous attendre à des variations fines d'ordre temporel : la hausse rapide de l'intensité entraînerait probablement un raccourcissement des latences accompagné d'une augmentation en amplitude pour les pics évoqués, et inversement pour la baisse rapide de l'intensité (Akhoun *et al.*, 2008a).

La hauteur de son est liée à la fréquence fondamentale F0, encodée de façon robuste par la FFR et donc facilement observable. C'est sans doute pourquoi cet indice phonologique a été au centre de la présente ligne de recherche.

En 2005, Krishnan a mis en évidence un effet de la langue maternelle sur l'encodage de la hauteur de son, en comparant les Speech ABR évoquées par des syllabes tonales pour deux groupes de participants chinois et américains. La FFR du premier groupe était alors plus robuste et suivait la trajectoire fréquentielle des syllabes tonales plus précisément que celle du second (Krishnan *et al.*, 2005), montrant alors une spécialisation de l'encodage neuronal en faveur des phonèmes de la langue maternelle.

Ce résultat a été précisé par la suite, en utilisant le même protocole à l'exception des *stimuli*. Cette fois-ci, la Speech ABR était enregistrée pour des trajectoires linéaires montante ou descendante, auquel cas l'avantage de la langue ne s'observait plus (Xu *et al.*, 2006) ; puis pour une trajectoire du mandarin ainsi que pour des trajectoires non rencontrées dans cette langue : l'effet de la langue maternelle ne s'observait alors que pour la trajectoire existante, n'influençant donc pas la capacité globale d'encodage des trajectoires fréquentielles, mais bien

l'encodage spécifique de phonèmes de la langue (Krishnan *et al.*, 2009a ; Krishnan & Gandour, 2009).

Notons que dans cette dernière expérimentation, les trajectoires du mandarin n'étaient pas portées par des syllabes du mandarin, mais par du bruit blanc (*iterated rippled noise* ; Yost, 1996). Aussi, une étude parue la même année a confirmé que l'effet de l'expertise linguistique sur la « Speech » ABR n'était pas spécifique au contexte, c'est-à-dire aux fréquences portées par la trajectoire (à son timbre, à son « habillage »), mais seulement à l'indice phonologique brut, en l'occurrence la trajectoire en soi (Krishnan *et al.*, 2009b). Cela ne signifie néanmoins pas que le timbre n'a pas d'effet sur la Speech ABR : une étude ultérieure a montré un encodage de F0 plus ample quand la trajectoire adoptait le timbre de la langue maternelle, contre le timbre d'une autre langue ou du bruit blanc (Krishnan *et al.*, 2011b).

La même équipe a montré par la suite que l'effet de l'expérience linguistique sur l'encodage neuronal des trajectoires s'étendait également aux accélérations de trajectoire non rencontrées en mandarin (Krishnan *et al.*, 2010a), aux syllabes d'autres langues tonales (nommément le thaï ; Krishnan *et al.*, 2010b), et à la dégradation de la structure temporelle fine des trajectoires (Krishnan *et al.*, 2010c).

En 2011, une première étude n'a pas retrouvé de bénéfice langue-dépendant pour des *stimuli* débarrassés de F0 (*missing fundamental*) ou d'harmoniques, appuyant de ce fait l'idée que ce bénéfice ne s'applique qu'aux phonèmes de la langue (Jeng *et al.*, 2011b). Aussi, une seconde étude a montré que l'effet langue-dépendant n'était pas présent à la naissance, ce qui suggère qu'il proviendrait bien de l'expérience linguistique (Jeng *et al.*, 2011d).

En changeant de paradigme pour l'étude d'individus bilingues, l'équipe de Kraus a quant à elle mis en évidence une augmentation de l'encodage spectral à F0, de la consistance de la Speech ABR à travers les essais, et des performances attentionnelles (Krizman *et al.*, 2012a ; Krizman *et al.*, 2014). L'interprétation des auteurs est que la pratique de deux langues bénéficie au système attentionnel, qui renforce alors l'encodage neuronal au niveau du tronc cérébral supérieur.

Tous ces résultats indiquent clairement l'existence d'effets de l'expertise linguistique sur l'encodage neuronal des sons de parole, et sont interprétés comme autant de signes d'une plasticité sous-corticale, dans le tronc cérébral supérieur d'où est supposée provenir la FFR. Notons que la réciproque a aussi été montrée (effet de l'encodage neuronal sur les capacités d'apprentissage d'une langue), avec un meilleur apprentissage des langues lié à un encodage plus robuste de la hauteur tonale dans le colliculus inférieur (Chandrasekaran *et al.*, 2012).

La thématique à laquelle nous allons nous intéresser maintenant est de bien loin la plus développée dans la littérature de la Speech ABR. Deux papiers princeps, parus en 2007, ont révélé que les musiciens avaient une *onset response* plus précoce, un encodage de la F0 plus ample, un meilleur suivi des variations de la hauteur tonale (Figure 3.17.), et une plus grande fidélité de la réponse au *stimulus* ; ces observations étaient de plus corrélées au nombre d'années de pratique instrumentale et à l'âge de son commencement, et étaient interprétées comme les conséquences d'une plasticité au niveau du tronc cérébral auditif (Wong *et al.*, 2007 ; Musacchia *et al.*, 2007).

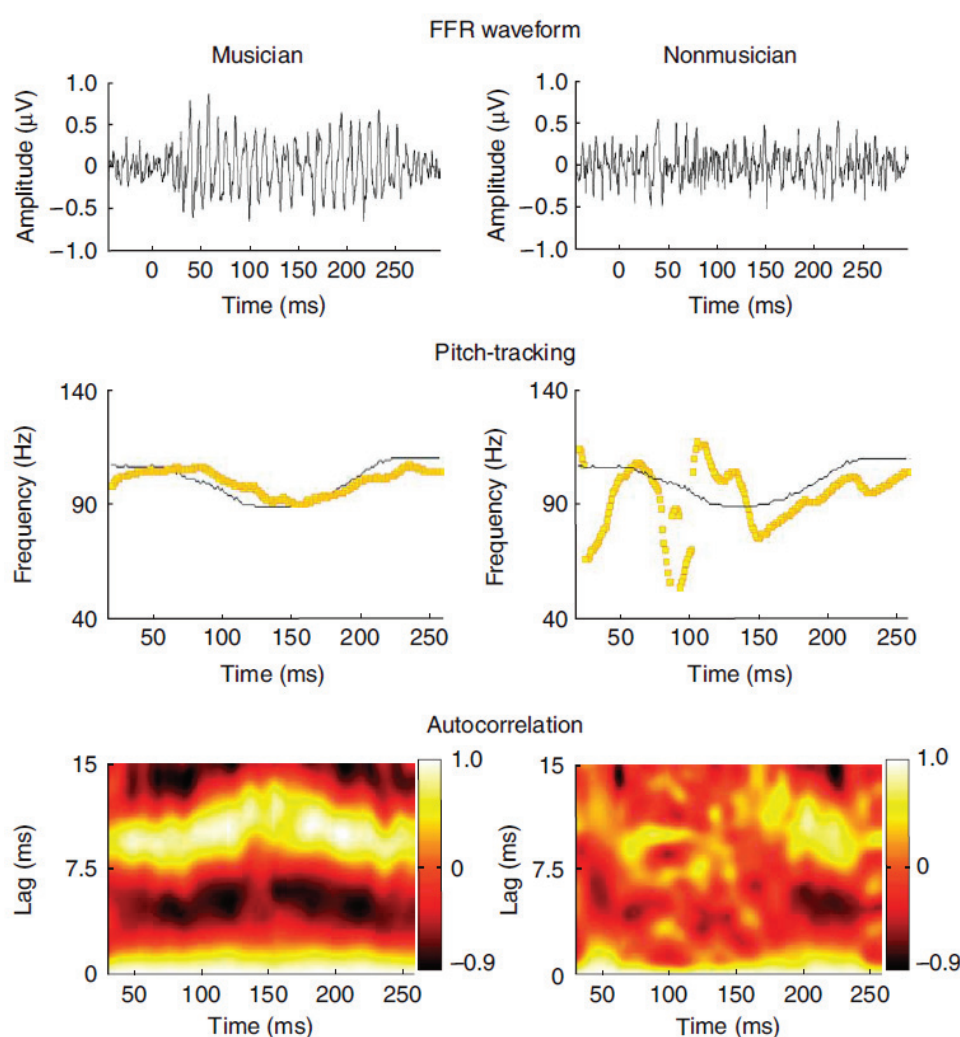


Figure 3.17. Meilleur suivi chez le musicien des variations de F0 d'une syllabe tonale /mi/ du mandarin (Wong *et al.*, 2007). Les données montrées sont individuelles. Au milieu, la ligne noire représente la trajectoire de F0 du *stimulus*, et le tracé orangé celle de la FFR ; les deux sont combinées dans l'étude pour calculer un coefficient de corrélation, indice de la qualité de suivi des variations de F0. En bas, le diagramme d'autocorrélation représente le coefficient d'autocorrélation en fonction de la période (inverse de F0) et du temps, pour la FFR seulement. Cette dernière mesure permet donc d'estimer la robustesse du suivi de phase, sans faire référence au *stimulus*.

Depuis 2007, un effort de recherche considérable a permis d'établir des liens entre pratique musicale et pratiquement toutes les thématiques abordées dans le présent chapitre :

- parole dans le bruit, avec une Speech ABR plus robuste au bruit chez les musiciens (Parbery-Clark *et al.*, 2009a,b, 2011a ; Bidelman & Krishnan, 2010 ; Chandrasekaran & Kraus, 2010a ; Parbery-Clark *et al.*, 2011c, 2013a ; Strait *et al.*, 2012b, 2013a) ;
- développement, avec un effet de la pratique musicale sur la Speech ABR pouvant se manifester dès l'âge de 3 ans, et perdurer longtemps après l'arrêt de la pratique (Strait *et al.*, 2013b ; Skoe & Kraus, 2013 ; Strait & Kraus, 2014) ;
- vieillissement, avec un retardement des déficits d'encodage temporel chez les musiciens, y compris pour une pratique musicale interrompue après l'enfance (Parbery-Clark *et al.*, 2011b, 2012b,c ; Skoe & Kraus, 2012 ; White-Schwotch *et al.*, 2013 ; Bidelman & Alain, 2015) ;
- expertise linguistique, avec des effets croisés conférant aux musiciens des améliorations dans l'encodage des sons de parole (Bidelman *et al.*, 2011a,b ; Krishnan *et al.*, 2012 ; Parbery-Clark *et al.*, 2012a ; Bidelman *et al.*, 2013a) ;
- troubles de l'audition, avec une Speech ABR plus robuste et une meilleure compréhension dans le bruit chez les musiciens (Parbery-Clark *et al.*, 2013b) ;
- perception catégorielle, avec une amélioration des performances comportementales accompagnée d'une Speech ABR plus ample au niveau du premier formant F1 (Bidelman *et al.*, 2014b) ;
- émotions, avec un encodage plus ample des parties complexes d'un son émotionnel (Strait *et al.*, 2009a,b) ;
- interactions audiovisuelles, avec un raccourcissement de latence et une augmentation en amplitude à l'*onset*, ainsi qu'un encodage plus ample à F0 lors de la présentation d'une vidéo de pratique instrumentale ou de prononciation de syllabe simultanément au *stimulus* acoustique, chez les musiciens davantage que chez les contrôles (Musacchia *et al.*, 2007).

L'idée globale ressortant de cette littérature abondante est que les musiciens ont un encodage neuronal des sons linguistiques ou musicaux bien plus fin et performant que les non-musiciens, et que la pratique musicale permet de « sculpter » les voies auditives nerveuses pour en améliorer le fonctionnement normal, la robustesse aux différents troubles liés à l'audition et la pérennité.

Selon les auteurs, cet affinage de l'encodage auditif sous-cortical pourrait être médié par l'action de voies corticofuges (*top-down*), l'activité corticale étant également améliorée par la pratique musicale (Musacchia *et al.*, 2008 ; Strait *et al.*, 2010 ; Strait & Kraus, 2011 ; Moreno & Bidelman, 2013).

Parallèlement aux études citées montrant les bénéfices d'une pratique musicale, des travaux ont considéré la perception de la musique de façon analytique, en liant la Speech ABR à différentes dimensions musicales, dont :

- le timbre (notion d'identité ou de « couleur » du son), avec une amélioration en amplitude de l'encodage spectral dans la FFR, d'autant plus lorsque celle-ci est évoquée par des sons de l'instrument pratiqué (Strait *et al.*, 2012a ; Kraus *et al.*, 2009) ;
- la consonance (caractère agréable d'un accord), avec une F0 plus ample dans la FFR à des accords harmoniques en comparaison à des accords inharmoniques ou dissonants, et un encodage plus robuste des accords dissonants chez les musiciens par rapport aux non-musiciens (Bidelman & Krishnan, 2009, 2011 ; Lee *et al.*, 2009 ; Bidelman *et al.*, 2011d ; Bidelman, 2013 ; Bones *et al.*, 2014) ;
- l'harmonie (caractère agréable d'une suite d'accords), avec un encodage spectral plus ample à F0 pour des accords précédés d'un accord harmoniquement (Marmel *et al.*, 2011) ;
- la gamme (série discrète de notes selon un schéma défini), avec une FFR plus ample au niveau de F0 pour des tons aux fréquences de la gamme diatonique comparativement à des tons hors de cette gamme, sur laquelle se base la musique occidentale (Bidelman *et al.*, 2011c)
- et le rythme (régularité temporelle des sons), avec un encodage plus robuste des sons présentés au tempo, donc réguliers et prédictibles (Tierney & Kraus, 2013a,b ; Strait *et al.*, 2011).

3.2.3.4. Déficiences auditives et troubles liés

Troubles de l'audition

En 2001, une première étude chez le malentendant a découvert un déficit dans le suivi du deuxième formant de la FFR spectrale (soustraction des réponses aux polarités opposées), pour des syllabes artificielles /ba/, /da/ et /ga/ avec F2 commençant respectivement à 900, 1600 et 2300 Hz (Plyler & Ananthanarayan, 2001). Onze ans plus tard, une nouvelle étude de Speech ABR met en évidence des stratégies d'encodage différentes chez un groupe d'adultes normo-entendants âgés entre 21 et 55 ans : les plus jeunes ont une meilleure représentation des basses fréquences, alors que les plus âgés se basent préférentiellement sur les hautes fréquences (Ruggles *et al.*, 2012). Cela expliquerait alors les plus grandes difficultés de compréhension dans le bruit rencontrées par les adultes normo-entendants (selon l'audiométrie tonale, donc) d'âge avancé, ces hautes fréquences étant particulièrement moins robustes au bruit que les basses fréquences (Moon *et al.*, 2014).

Plusieurs études ont depuis précisé l'impact des troubles de l'audition sur la Speech ABR. Assez paradoxalement, une surreprésentation des basses fréquences a été mise en évidence chez

les patients malentendants (Figure 3.18.), interprétée comme un déséquilibre en défaveur des hautes fréquences nécessaires à la compréhension des phonèmes (Anderson *et al.*, 2013a). Paradoxalement car dans ce dernier cas, la représentation préférentielle des basses fréquences ne confère pas d'avantage pour la compréhension dans le bruit, comme décrit plus haut ; cela pourrait provenir d'un masquage énergétique, d'un recouvrement de l'activité électrophysiologique des hautes fréquences par celle des basses fréquences (Kraus & Anderson, 2014).

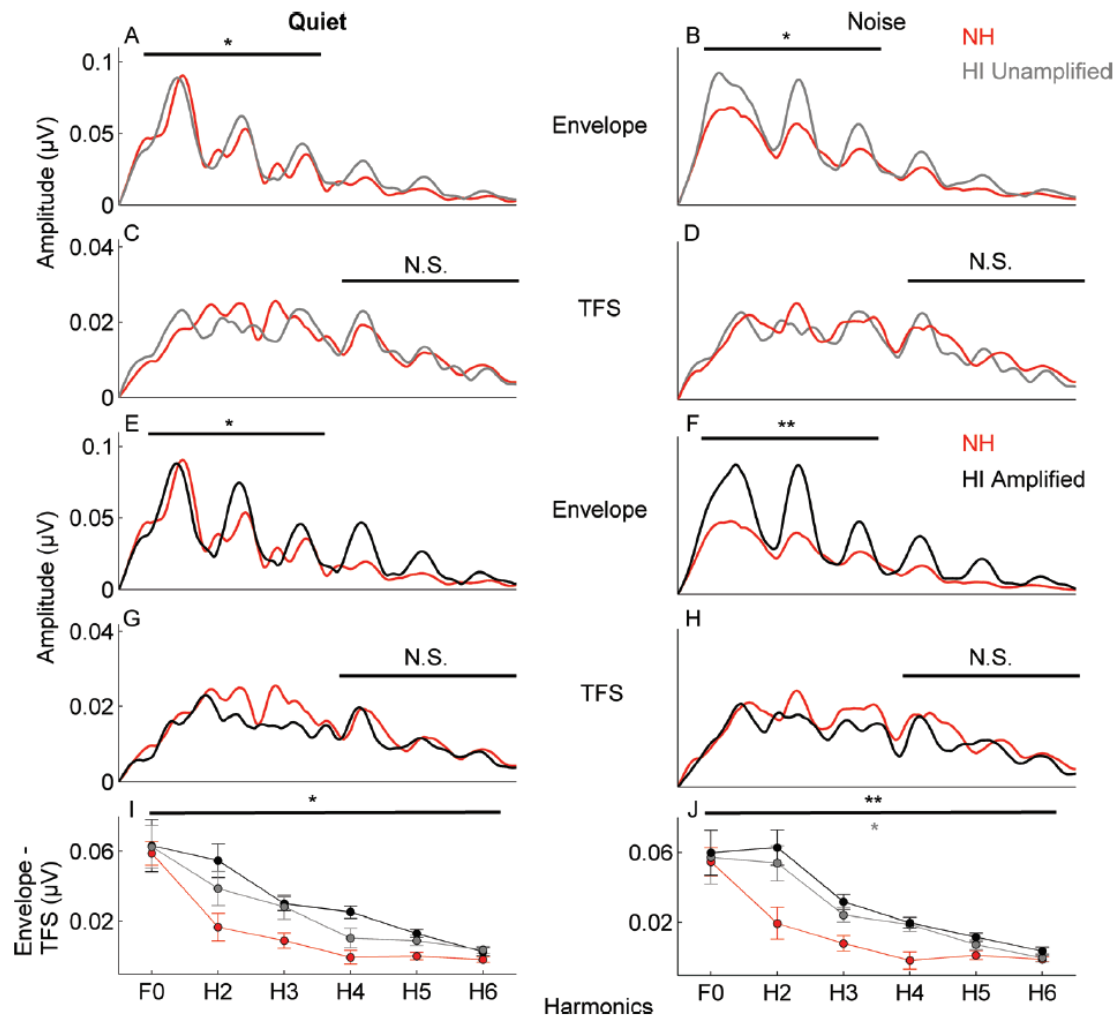


Figure 3.18. Surreprésentation des basses fréquences chez le patient malentendant (Anderson *et al.*, 2013a). Les spectres de la FFR sont représentés pour la syllabe /da/ de 40 ms délivrée dans le calme (colonne de gauche) ou dans le bruit (colonne de droite), et qui peut être amplifiée ou non pour les malentendants (HI, *hearing impaired*). L'amplification s'effectuait directement sur le *stimulus* afin de compenser la perte auditive, soit sur un fichier son propre à chaque patient testé. Enfin, la FFR d'enveloppe et la FFR spectrale (TFS, *temporal fine structure*) étaient étudiées séparément, et combinées dans la ligne du bas pour avoir une estimation de leur balance à F0 (d'environ 110 Hz) et ses harmoniques (H6 étant donc à 660 Hz, pour un premier formant du /da/ à 720 Hz). La surreprésentation des basses fréquences ne s'observe logiquement qu'au niveau de la FFR d'enveloppe, et davantage dans le bruit et dans la condition amplifiée des malentendants. Dans la ligne du bas, un effet significatif du groupe s'observe sur toutes les fréquences ; dans le bruit seulement, un effet significatif de l'amplification est observé également sur toutes les fréquences.

Enfin, une dernière étude a montré une différence entre malentendants musiciens et non-musiciens : les premiers ont une fréquence fondamentale plus robuste et une meilleure compréhension dans le bruit, mais pas d'augmentation dans l'encodage des hautes fréquences (Parbery-Clark *et al.*, 2013). Cependant, les Speech ABR des malentendants musiciens présentent une plus grande précision telle que mise en évidence par la corrélation au *stimulus*, ainsi que des latences plus précoces pour les réponses transitoires.

L'ensemble de ces résultats ne permet pas de conclure fermement sur le rôle de l'encodage de F0 pour les performances de compréhension dans le bruit : bénéficiant aux jeunes adultes normo-entendants et aux malentendants musiciens, la surreprésentation de F0 marque un déficit chez les malentendants non-musiciens. Ainsi, c'est peut-être plus la précision temporelle que la robustesse spectrale qui conférerait une plus grande compréhension dans le bruit. De futures études sont nécessaires pour lever cette ambiguïté.

Appareillage auditif

Aucune étude n'a à ce jour montré de Speech ABR chez le patient malentendant appareillé. Cependant, ce marqueur est un excellent candidat pour objectiver finement l'encodage spectro-temporelle des différentes informations contenues dans les sons de parole, et pourrait alors remplir diverses rôles dans l'appareillage des patients malentendants. Deux études ont anticipé quels pouvaient être ces rôles : caractérisation précise des troubles de l'audition, prédiction de succès de l'appareillage en fonction des capacités d'encodage temporel, calcul optimal des paramètres de réglage des aides auditives en se basant sur les déficits observés dans la Speech ABR, ou encore suivi des bénéfices apportées par les aides auditives en comparant l'évolution de ce marqueur (Dajani *et al.*, 2013 ; Anderson & Kraus, 2013).

Troubles des apprentissages

Cette thématique a été l'objet de nombreuses études depuis la découverte en 2001 d'un déficit d'encodage spectro-temporel de la parole dans le bruit chez les enfants atteints de troubles des apprentissages (Cunningham *et al.*, 2001). Ces études ont mis en évidence des retards de latence aux composantes transitoires et soutenues de la Speech ABR (Kraus 2001 ; King *et al.*, 2002 ; Kouni *et al.*, 2006 ; Rocha-Muniz *et al.*, 2012 ; Kouni *et al.*, 2013 ; Malayeri *et al.*, 2014), un encodage spectral dégradé au niveau du premier formant (Wible *et al.*, 2004 ; Hornickel *et al.*, 2012a), une défaillance du suivi des trajectoires fréquentielles (Basu *et al.*, 2010), une plus grande variabilité des essais unitaires (Hornickel & Kraus, 2013), ainsi qu'une altération de l'intégration audiovisuelle (Hayes *et al.*, 2003a). Ces déficits observés dans une grande proportion d'enfants atteints de troubles des apprentissages (Banai *et al.*, 2007) sont par ailleurs liés aux déficits observés au niveau cortical (Wible *et al.*, 2005 ; Song *et al.*, 2008a).

L'avancée des connaissances sur le sujet a permis d'associer Speech ABR et capacités de lecture (Banai *et al.*, 2009). Notamment, une imprécision temporelle dans la réponse aux transitions formantiques régissant la perception des consonnes est liée à des déficits de lecture (Johnson *et al.*, 2007 ; Hornickel *et al.*, 2009b), si bien que la qualité d'encodage temporel de la Speech ABR peut être utilisée pour prédire les capacités de lecture (Hornickel *et al.*, 2011).

Des stratégies de remédiation par entraînement auditif ou d'utilisation d'un système FM en salle de classe (voix de l'instituteur captée par un microphone et transmise par ondes radio aux écouteurs de l'enfant) ont montré leurs effets bénéfiques sur la compensation des troubles des apprentissages, ainsi que sur la qualité des Speech ABR sous-jacentes (King *et al.*, 2002 ; Hornickel *et al.*, 2012c ; Figure 3.19.).

Soulignons enfin un intérêt notable qu'a ici la Speech ABR : le diagnostic de la dyslexie se pose actuellement sur le constat de deux ans de retard sur l'âge de lecture. Or, ce retard est très difficile à rattraper par la suite, mettant alors l'enfant en difficulté scolaire dès le début de son cursus. Grâce à son rôle prédictif, la Speech ABR peut être utilisée pour détecter la dyslexie avant que les déficits comportementaux ne se manifestent.

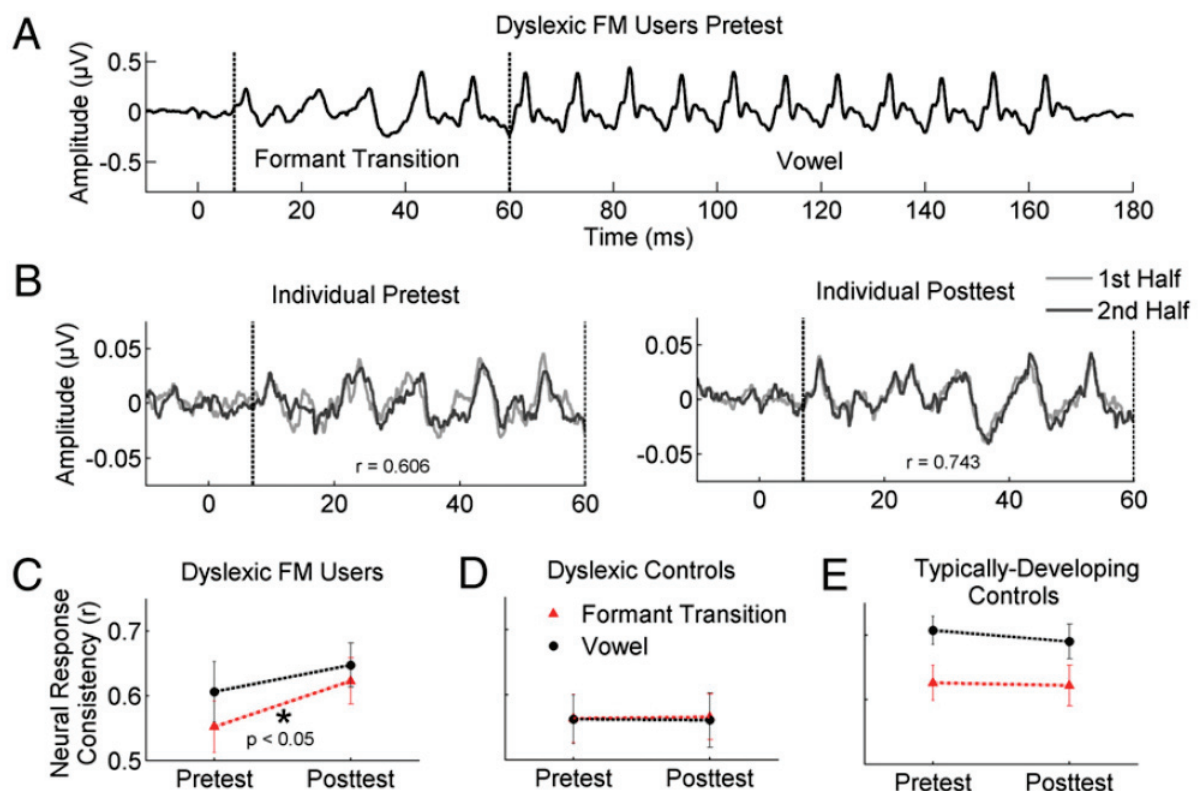


Figure 3.19. Amélioration en stabilité de la représentation neuronale des sons de parole suite à l'utilisation d'un système FM (Hornickel *et al.*, 2012c). Cette stabilité, aussi appelée consistance ou reproductibilité, correspond au coefficient de corrélation de deux sous-moyennes des essais individuels, dans le cas présent entre les essais de la première moitié de la session d'enregistrement et ceux de la seconde moitié. Une amélioration significative n'est observée que chez les enfants dyslexiques équipés des prothèses FM, et seulement pour la partie de la réponse représentant la transition formantique.

3.2.3.5. Cognition

Perception catégorielle

Dans la perception catégorielle, des syllabes d'un *continuum* (par exemple entre /ba/ et /da/) ne sont pas perçues différentes à moins d'être situées de part et d'autre d'une frontière catégorielle séparant les deux phonèmes (/ba/ et /da/ dans l'exemple ; Liberman *et al.*, 1957). Autrement dit, la perception est moins fidèle à l'acoustique de chaque syllabe, qu'aux catégories phonémiques apprises ; pour reprendre la terminologie du point précédent, elle est plus intrinsèque qu'extrinsèque.

Une étude récente a exploré les liens entre perception catégorielle et Speech ABR, sans mettre en évidence un quelconque effet de la perception sur l'encodage spectro-temporel des *stimuli*, ainsi considéré purement extrinsèque (Bidelman *et al.*, 2013b).

Effet de transformation verbale

La deuxième étude ayant été recensée comme utilisant la Speech ABR a investigué le lien entre percept et encodage neuronal (Galbraith *et al.*, 1997). Dans l'effet de transformation verbale, une syllabe donnée et répétée régulièrement un grand nombre de fois va donner lieu à une perception modifiée, en faveur d'une syllabe phonétiquement proche (Warren, 1968). Au niveau enregistré par la Speech ABR, la question se pose alors quant à l'information représentée : acoustique et donc invariable, ou perceptive et donc sujette à transformation ; l'encodage sera alors dit extrinsèque ou intrinsèque, respectivement.

En divisant le groupe de participants entre ceux manifestant peu l'effet de transformation verbale, et ceux y étant particulièrement sensibles, Galbraith a montré une modification en amplitude de la FFR évoquée par les essais perçus de façon transformée (augmentée chez les participants peu sensibles, et diminuée chez les participants sensibles), interprétée comme un effet des voies descendantes perceptives et attentionnelles.

Attention sélective

Comme nous l'avons déjà décrit lors de l'introduction de cette présente partie, une des premières études de Speech ABR a montré un effet de l'attention sur la FFR : dans une tâche d'écoute dichotique présentant à chaque oreille une voyelle d'un locuteur différent (/e/ prononcé par une voix d'homme avec une F0 de 170 Hz, et /a/ prononcé par une voix de femme avec une F0 de 230 Hz), l'amplitude de F0 était plus grande lorsque l'attention était portée sur la voyelle donnée (Galbraith *et al.*, 1998). Ce résultat, par ailleurs reproduit dans une plus récente étude (Lehmann & Schönwiesner, 2014), suggère une modulation de l'encodage spectral par les voies descendantes du système attentionnel. Aussi, il peut être mis

en relation avec une précédente étude ayant montré un effet attentionnel similaire sur l'amplitude des OAE (Giard *et al.*, 1994) : l'information spectrale pourrait donc être amplifiée par l'attention dès le niveau cochléaire, se traduisant par un signal plus fort dans les voies auditives ascendantes, et expliquant alors l'effet observé sur la Speech ABR.

Emotions

Dans une étude présentant des images émotionnellement chargées à des participants pendant l'enregistrement de Speech ABR, aucun effet des émotions n'a pu être mis en évidence dans cette réponse (Wang *et al.*, 2010).

Interactions audiovisuelles

La présentation visuelle d'un locuteur prononçant une syllabe a été montrée à même de rallonger la latence d'un pic dans l'*onset response* de la Speech ABR (Musacchia *et al.*, 2006). Une seconde étude a révélé un effet audiovisuel résultant en une augmentation de l'amplitude de la F0 dans le spectre de la FFR, d'autant plus importante chez des participants musiciens (Musacchia *et al.*, 2007). Ces deux résultats montrent la sensibilité multimodale de la Speech ABR, qui peut être modifiée dès 10 ms après le début du *stimulus* auditif.

3.2.3.6. Pathologies

Nous allons ici passer rapidement en revue les quelques études ayant questionné les altérations éventuelles de la Speech ABR dans des pathologies neurologiques ou psychiatriques, études cependant peu nombreuses en regard de la production bibliographique abordant les thématiques du langage, de la musique, de la perception dans le bruit ou des troubles des apprentissages.

Amusie congénitale

Dans cette pathologie neurologique affectant la cognition de l'information musicale (Albouy *et al.*, 2013), une étude récente a mis en évidence une dégradation de la précision temporelle et de l'encodage des hautes fréquences dans la Speech ABR de patients amusiques (Lehmann *et al.*, 2015), après une première tentative qui était restée vaine (Liu *et al.*, 2014).

Cécité

Les patients non-voyants congénitaux présentent des latences raccourcies aux pics des réponses *onset* et *offset*, ainsi qu'un encodage spectral plus robuste sur toutes les fréquences, par rapport à des individus voyants (Jafari & Malayeri, 2014), ce qui peut être dû à des effets attentionnels, portés sur la modalité auditive à défaut de disposer d'informations visuelles.

Epilepsie

A ce jour, une seule étude a montré une latence des pics de l'*onset response* retardée chez des enfants épileptiques, d'autant plus que leur âge est bas, sans qu'il n'y ait d'effet du groupe sur la latence de l'onde V des Click ABR (Elkabariti *et al.*, 2014). Ce résultat suggère un déficit dans l'encodage temporel des indices transitoires, sans que l'étude n'ait pu mettre en évidence de déficit dans l'encodage spectral (FFR).

Hyperactivité

Chez des enfants atteints de troubles d'hyperactivité et de déficits attentionnels (ADHD), une étude a montré un rallongement des latences des pics III et V des Click ABR ainsi que de la plupart des pics de la Speech ABR, sans mettre en évidence de déficit d'encodage spectral au niveau de la FFR (Jafari *et al.*, 2015).

Schizophrénie

Un article récent a proposé l'utilisation de la Speech ABR dans la caractérisation des troubles auditifs liés à la psychose schizophrénique (Tarasenko *et al.*, 2014), des liens entre colliculus inférieur et hallucinations auditives ayant été précédemment établis par les Click ABR (Lindstrom *et al.*, 1987) et l'IRM fonctionnelle (Shergill *et al.*, 2000).

Troubles du spectre autistique

Chez l'enfant atteint de troubles du spectre autistique, deux études ont montré un déficit dans le suivi de F0 (Russo *et al.*, 2008), et dans la latence et l'amplitude des différents pics de la Speech ABR pour des syllabes présentées dans le calme ou dans le bruit, avec une dégradation encore plus importante dans le bruit (Russo *et al.*, 2009). L'incapacité de suivre fidèlement les variations de F0 pourraient sous-tendre les problèmes de perception de la prosodie, liée à l'interprétation des émotions véhiculées par la voix humaine, entre autres.

De façon intéressante, un entraînement auditif permet chez certains enfants autistes d'améliorer ces déficits spectro-temporels (Russo *et al.*, 2010). L'entraînement en question durait de cinq à dix semaines, et consistait en une série de sept jeux d'ordinateur incluant des leçons d'écoute et des tâches d'attention auditive, de discrimination et de mémoire (*Fast ForWord*, Scientific Learning).

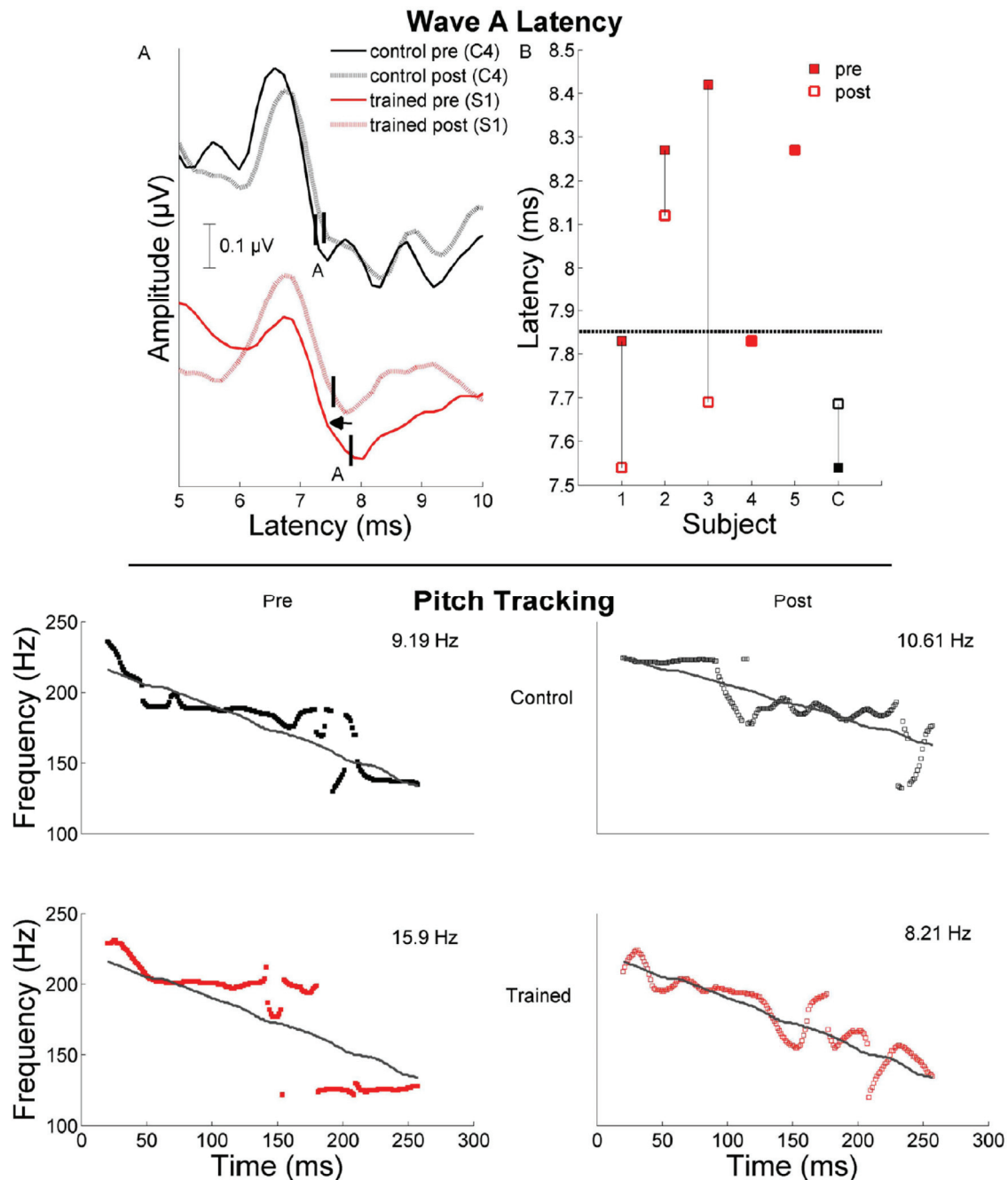


Figure 3.20. Amélioration des déficits spectro-temporels d'encodage des sons de parole chez l'enfant avec autisme après entraînement auditif (Russo *et al.*, 2010). Le groupe contrôle est composé de 6 enfants avec autisme n'ayant pas effectué l'entraînement. Dans la partie haute, la latence du pic transitoire A (rebond négatif du pic V de l'*onset response*) est raccourcie (à gauche, données individuelles), et cet effet s'observe chez 3 des 5 enfants avec autisme de l'étude (à droite ; les points noirs représentent la moyenne du groupe contrôle). Dans la partie basse, le suivi de la F0 est amélioré uniquement chez les enfants ayant effectué l'entraînement.

3.2.3.7. Connaissances générales sur la Speech ABR

Effet du genre

Deux études ont montré une amélioration de la Speech ABR en certains aspects chez les femmes : des latences plus précoces et une amplitude plus élevée s'observent au niveau de l'*onset response*, mais pas dans la FFR ; aussi, les fréquences correspondant à F0 et au premier formant F1 sont plus amples (Krizman *et al.*, 2012b ; Ahadi *et al.*, 2014a). Pour les auteurs, cela peut expliquer une plus grande prévalence des troubles du langage pour le genre masculin.

Facteurs génétiques et environnementaux

Les Speech ABR des membres d'une fratrie se ressemblent davantage que celles d'enfants sans lien de parenté, quand bien même appariés par niveau de lecture. Cette comparaison permet de réduire la probabilité qu'il s'agisse d'un effet de l'environnement familial et éducatif, et suggère donc un facteur génétique sur l'encodage des sons de parole (Hornickel *et al.*, 2013).

Lien à d'autres mesures (MMN, OAE)

A l'aide d'un paradigme *oddball* (consistant en une séquence composée d'une majorité de *stimuli* identiques, dits « standards », parsemée de rares *stimuli* différents, dits « déviants »), une équipe a montré une corrélation entre une latence anormale des pics de l'*onset response* et une dégradation de la MMN (*Mismatch negativity*) corticale (Banai *et al.*, 2005). Aussi, une autre équipe a mis en évidence une réduction en amplitude des harmoniques 2 et 4 (F0 de 100 Hz) dans la FFR évoquée par une syllabe déviante (Slabu *et al.*, 2012). Ces résultats montrent l'existence d'interactions entre les générateurs de la Speech ABR et ceux de la MMN corticale, sous-tendant la détection d'irrégularités dans une séquence de *stimuli*. Par ailleurs, et simplement pour la citer puisqu'elle est écrite en chinois, une étude a reproduit ces résultats de corrélations entre latences et amplitudes de la Speech ABR et de la MMN (Fu *et al.*, 2015).

A l'échelle cochléaire, des études se sont intéressées au lien entre otoémissions acoustiques (*Otoacoustic emissions, OAE*) et Speech ABR. Deux études ont ainsi mis en évidence une corrélation entre la robustesse des OAE d'une part, et les latences des composantes transitoires ou la robustesse d'encodage de la Speech ABR d'autre part (Dhar *et al.*, 2009 ; Rana & Barman, 2011). Il n'est en effet pas inconcevable que des cellules ciliées externes nombreuses, intègres et fonctionnelles (et donc engendrant de fortes otoémissions) puissent participer à une transmission plus fidèle du signal acoustique au nerf auditif par le biais des cellules ciliées internes, et ainsi entraîner le recrutement et la synchronisation d'un plus grand nombre de fibres dans les voies auditives ascendantes, et par conséquent une *onset response* plus précoce. Le lien entre fonction cochléaire et Speech ABR a été par ailleurs étudié, montrant un effet des filtres cochléaires sur la latence et l'amplitude de la Speech ABR évoquée par les différentes

octaves d'un *stimulus* filtré : la latence de cette réponse diffère en fonction de la bande de fréquence délivrée, et la FFR est maximale pour le *stimulus* filtré entre 2 et 4 kHz (Nuttall *et al.*, 2015), rappelant ainsi la sensibilité fréquentielle des Click ABR (Stapells & Oates, 1997).

3.2.4. Etat des lieux

Comme nous avons pu le constater dans l'inventaire des études de Speech ABR que nous venons de parcourir, de nombreux pas ont été faits depuis le premier enregistrement de cette réponse, il y a 20 ans.

Nous savons désormais que la Speech ABR est altérée dans un nombre conséquent de pathologies sensorielles, neurologiques ou psychiatriques, et qu'elle a un rôle de prédicteur pour les troubles de l'audition dans le bruit, ainsi que pour les troubles des apprentissages de type dyslexie. Ce dernier résultat constitue d'ailleurs une avancée majeure dans la prise en charge clinique potentielle des enfants dyslexiques, en permettant de détecter les troubles avant qu'ils ne s'installent plus durablement et ne produisent leurs effets néfastes sur l'apprentissage de la lecture.

Nous savons aussi que la Speech ABR n'est pas figée mais bien au contraire hautement dynamique, puisqu'elle montre une plasticité à court (en minutes ou heures, par exemple avec l'effet du contexte), moyen (en mois, par exemple avec les entraînements auditifs) et long terme (en années, par exemple avec l'apprentissage d'une langue ou d'un instrument de musique). Or, ce type d'activité bas-niveau, c'est-à-dire encodant fidèlement les détails du *stimulus*, était vu avant ces résultats comme un processus statique de transmission du signal entre la cochlée et le cortex auditif.

Nous savons encore que la Speech ABR s'articule avec la fonction cochléaire (par les OAE, reflétant aussi l'activité du système efférent olivo-cochléaire médian, soient des premiers relais sous-corticaux), l'activité corticale (complexe P50-N1-P2, réponse MMN) et certains processus cognitifs, attentionnels et multi-sensoriels (effet de transformation verbale, attention sélective dans une tâche d'écoute dichotique, interactions audiovisuelles).

Nous savons enfin que la Speech ABR représente fidèlement les composantes spectro-temporelles des consonnes occlusives, des voyelles et des syllabes tonales.

Cependant, de nombreuses zones d'ombres restent à explorer pour les années à venir.

Nous n'avons par exemple aucune information quant à l'encodage d'autres types de sons de parole, tels que par exemple les consonnes fricatives (/f/, /v/, /s/, /z/, pour en citer

quelques-unes). Or, leur composition haute-fréquence les rend particulièrement sensibles au bruit, ainsi qu'aux pertes auditives telles que la presbyacousie.

Jusqu'à présent, la méthodologie de la Speech ABR est restée relativement semblable à celle des débuts, avec le recueil quasi-systématique d'un nombre prédéfini d'essais sur un nombre limité d'électrodes, suivi d'une analyse des latences et amplitudes dans le domaine temporel, d'une mesure du niveau d'énergie des différentes composantes spectrales, et plus rarement d'analyses de corrélation pour déterminer la consistance de la réponse ou son suivi de F0. Plusieurs pistes pourraient pourtant débloquent certaines approches ou mesures. Par exemple, l'utilisation de protocoles adaptatifs pourrait réduire la durée d'enregistrement, en arrêtant le recueil aussitôt que la réponse aurait atteint le niveau de qualité souhaité ; il serait alors possible d'utiliser une plus grande variété de *stimuli*, comme des mots voire des phrases. Aussi, l'utilisation de systèmes EEG dotés d'un plus grand nombre de voies permettrait également de diminuer le temps d'acquisition, en profitant des techniques multi-électrodes de débruitage pour obtenir une Speech ABR de qualité souhaitée en un moindre nombre d'essais moyennés ; de plus, cette technique offrirait un accès à de nombreuses analyses supplémentaires telles que l'étude des topographies ou leur utilisation dans la reconstruction de sources.

Au niveau théorique, les résultats concernant les malentendants sont relativement contradictoires, avec un rôle de l'encodage de F0 bénéfique pour les jeunes normo-entendants et les musiciens âgés malentendants, mais délétère pour les malentendants âgés non-musiciens : de futures études seraient souhaitables pour clarifier ce point, d'autant plus lorsque l'on sait la prévalence croissante des troubles de l'audition. Par ailleurs, aucune étude n'a étudié l'effet d'un appareillage auditif sur l'encodage spectro-temporel des sons de parole, malgré l'attrait de cette thérapeutique permettant bien souvent de redonner un confort d'écoute aux patients.

Enfin, les générateurs neuronaux de la Speech ABR demeurent largement méconnus, bien qu'ils soient communément admis comme étant sous-corticaux. Ainsi, la majorité des études que nous avons présentées dans ce chapitre interprètent tout changement de la Speech ABR comme une preuve de plasticité du tronc cérébral auditif. Or, si les réponses transitoires semblent difficilement pouvoir être générées au niveau cortical, des résultats d'un autre domaine de recherche (notamment l'étude anatomo-fonctionnelle du cortex auditif en utilisant l'EEG intracrânien) ont montré l'existence d'une réponse corticale verrouillée en phase avec le *stimulus*, et ce jusqu'à une F0 de 114 Hz. L'hypothèse d'une participation corticale à la FFR enregistrée au niveau du scalp n'est donc pas à exclure et pourrait, si elle se confirme, changer la façon dont nous voyons actuellement la Speech ABR.

Deuxième partie

Travaux expérimentaux

4. Axe « Générateurs neuronaux de la Speech ABR »

Comme nous l'avons vu dans la partie 3.2.2, la connaissance des générateurs de la Speech ABR se base en très grande partie sur des résultats concernant des réponses à des sons plus élémentaires (clics, tons purs ou complexes). Or, comme nous avons pu le préciser à plusieurs endroits dans la partie introductive, le système auditif humain est non-linéaire, c'est-à-dire qu'on ne peut prédire sa réponse à une structure sonore complexe par une simple combinaison des réponses aux parties isolées de cette structure. Aussi, l'hypothèse sous-corticale se base sur nombre d'idées préconçues à propos des capacités de synchronie de phase (*phase-locking*) et des latences d'arrivée de l'information auditive dans les différents relais des voies auditives. Il nous semblait donc légitime de remettre cette hypothèse en question, potentiellement au grand dam de quantité d'études l'ayant utilisée pour l'interprétation de leurs résultats.

Pour aborder la question des générateurs neuronaux de la Speech ABR, nous avons d'abord développé une méthodologie de recueil topographique simultané de la Speech ABR et des PEA corticaux. Dans un second temps, et en se basant sur l'observation des topographies apportées par notre méthodologie innovante, l'activité du cortex auditif humain a été enregistrée *in situ* en SEEG.

Ces deux études ont donné lieu à deux publications, la première publiée et la seconde en relecture avant soumission, qui sont ici reproduites en l'état.

4.1. Topographie de la Speech ABR

L'article intitulé [Enregistrements topographiques des potentiels évoqués auditifs en réponse à la parole : Réponses sous-corticale et corticales] (*Topographic recordings of auditory evoked potentials to speech : Subcortical and cortical responses*) a été publié en 2015 dans le journal à comité de lecture *Psychophysiology*, numéro 52, pages 594 à 599.

BRIEF REPORT

Topographic recordings of auditory evoked potentials to speech: Subcortical and cortical responses

LUDOVIC BELLIER,^{a,b} PATRICK BOUCHET,^{a,b} ARNAUD JEANVOINE,^{a,b} OLIVIER VALENTIN,^{a,b}
HUNG THAI-VAN,^{a,b,c} AND ANNE CACLIN^{a,b}

^aLyon Neuroscience Research Center, Brain Dynamics and Cognition team, INSERM U1028, CNRS UMR5292, Lyon, France

^bUniversity Lyon 1, Lyon, France

^cHospices Civils de Lyon, Service d'Audiologie et Explorations Orofaciales, Hôpital Edouard Herriot, Lyon, France

Abstract

Topographies of speech auditory brainstem response (speech ABR), a fine electrophysiological marker of speech encoding, have never been described. Yet, they could provide useful information to assess speech ABR generators and better characterize populations of interest (e.g., musicians, dyslexics). We present here a novel methodology of topographic speech ABR recording, using a 32-channel low sampling rate (5 kHz) EEG system. Quality of speech ABRs obtained with this conventional multichannel EEG system were compared to that of signals simultaneously recorded with a high sampling rate (13.3 kHz) EEG system. Correlations between speech ABRs recorded with the two systems revealed highly similar signals, without any significant difference between their signal-to-noise ratios (SNRs). Moreover, an advanced denoising method for multichannel data (denoising source separation) significantly improved SNR and allowed topography of speech ABR to be recovered.

Descriptors: Speech ABR, FFR, N1, Event-related potential, Brainstem

Speech auditory brainstem response (speech ABR) is a scalp-recorded, high frequency event-related potential (ERP) evoked by speech stimuli. As speech contains brief transitions and long periodical parts, speech ABR has two types of components: transient and sustained, respectively identified as onset/offset response and frequency-following response (FFR). With its high specificity to each speech stimulus, speech ABR is an ideal electrophysiological marker of speech encoding, and has become widely used in numerous studies assessing a large range of hearing particularities, from expertise to impairments. For instance, musicians' ability to better hear speech in noise is related to their faster neural timing as revealed by speech ABR (Parbery-Clark, Strait, Hittner, & Kraus, 2013). And poor pitch tracking in speech ABR's FFR component has been related to the prosody-encoding deficits of children with autism spectrum disorders (Russo et al., 2008). Speech ABR also represents one of the most promising tools for hearing impairment diagnosis and management, by offering exhaustive

spectrotemporal information about each type of hearing loss (Anderson & Kraus, 2013; Dajani, Heffernan, & Giguere, 2013).

When it comes to speech ABR recording, clinical electroencephalography (EEG) systems originally designed for click ABR measurements are most often used, as they allow for the recording of high frequency activities with fine temporal resolution and decent signal-to-noise ratios due respectively to their high sampling rate (7 to 50 kHz) and amplifier quality. However, such EEG systems have a major limit: their very low number of channels (often three including ground and reference), which constrains the experimenter to use the classical "vertical montage" (vertex referenced at the mastoid ipsilateral to the stimulated ear or at linked mastoids). As a result, the experimenter has no access to the topographical information required, for example, for speech ABR generator(s) characterization, a topic scarcely debated in the literature. Access to speech ABR topography could also allow the enhancement of the description of populations of interest.

Obtaining speech ABR topographies could be done using research EEG systems, which are classically used to record cortical auditory evoked potentials (AEPs) on a large number of channels (32 to 256), at the expense of a relatively low sampling rate (below 5 kHz). In principle, referring to the Nyquist-Shannon sampling theorem (the maximum observable frequency is half the sampling frequency) and to the brainstem phase-locking limit of 1500 Hz (reviews in Galbraith et al., 2000; Richard, Jeanvoine, Veuillet, Moulin, & Thai-Van, 2010), such a system should be able to record proper speech ABR. Moreover, almost every speech ABR study

The authors are grateful to Alain de Cheveigné for his advice on how to use DSS. This work was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) and the LABEX CELYA (ANR-11-LABX-0060) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). LB was funded by a Ph.D. fellowship from the Région Rhône Alpes.

Address correspondence to: Ludovic Bellier, Lyon Neuroscience Research Center, Brain Dynamics and Cognition team, 69675 Bron cedex, France. E-mail: ludovic.bellier@inserm.fr

uses alternating polarities for stimulus presentation. This technique consists in presenting the stimulus and its opposite phase in equal amounts, then averaging speech ABRs to both polarities: this can minimize a possible stimulus artifact and the cochlear microphonics contribution to the response, yet it also emphasizes lower frequencies up to 800 Hz (Aiken & Picton, 2008). A tutorial on speech ABR recordings (Skoe & Kraus, 2010) reports that it is possible to use low sampling rate EEG systems (~6 kHz) for speech ABR recording. A few studies have indeed reported speech ABRs with sampling frequencies below 5 kHz (Aiken & Picton, 2006; Dajani, Purcell, Wong, Kunov, & Picton, 2005; Prévost, Laroche, Marcoux, & Dajani, 2013). However, none of them have compared the resulting speech ABRs with those recorded with higher sampling rate EEG systems. Controlling the impact of using a conventional research low sampling rate, multichannel EEG system on speech ABR quality is thus a mandatory step prior to further processing and analysis of such data.

To this aim, we first compared speech ABRs simultaneously recorded with two EEG systems: a 4-channel clinical EEG system with a sampling rate of 13.3 kHz, and a 32-channel research EEG system with a sampling rate of 5 kHz. Then, using data from the research EEG system only, we obtained concomitant cortical AEPs and their topographies and, for the first time to our knowledge, the topography of speech ABR, thanks to an advanced method of signal processing based on spatial denoising, allowed by the

32-channel recordings. We also show how multichannel recordings can be used to reduce acquisition time.

Method

Participants

Fourteen healthy adults (ages 20–48 years, $M = 28$ years; 5 women) participated in the study. All were normal-hearing (no threshold below 20 dB HL between 250 and 8000 Hz in tonal audiometry, and wave V latencies below 6.4 ms in click ABRs) native French speakers, and none had known language, psychiatric, or neurologic impairment. Each participant gave his/her informed written consent. All experimental procedures were carried out following the Declaration of Helsinki and with the agreement of the Ethics Committee from Lyon's Leon Bérard Center (CCPPRB #05-024).

Stimulus

Speech stimulus was a 200-ms natural /ba/ syllable, recorded from a French female voice (Figure 1a). It was composed of a 108-ms voicing followed by a 92-ms vowel. Fundamental frequency of the vowel /a/ varied slightly around 200 Hz, and the four first formants were respectively centered around 825, 1490, 3070, and 4430 Hz. A 30-ms silence was inserted at the beginning of the file to enable the recording of a baseline epoch in the EEG data (see below).

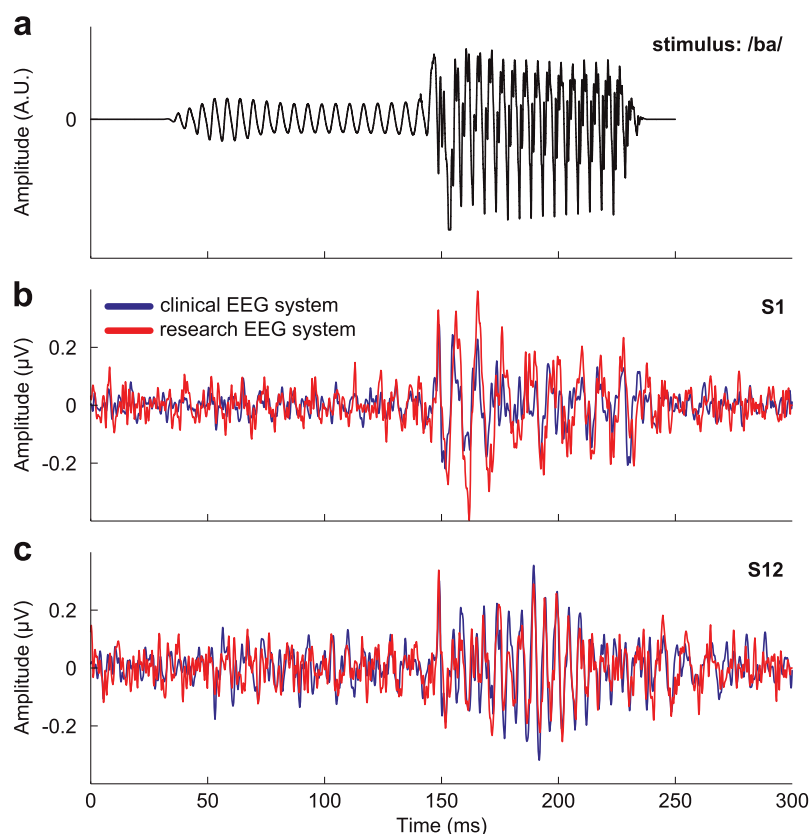


Figure 1. Speech ABR evoked by a 200 ms-long /ba/ syllable (a; A.U. = arbitrary units), simultaneously recorded with clinical and research EEG systems, for two representative participants (b,c). For both participants, speech ABR from clinical and research EEG systems highly resemble each other, with similar noise level (0 to 30 ms), signal amplitude, and time course (150 to 225 ms). Note the FFR difference between S1 (b) and S12 (c), respectively, following 100 Hz (/a/ fundamental frequency subharmonic) and 200 Hz.

Procedure

Participants sat in an armchair in a quiet room. The stimulus was binaurally delivered through electrically shielded insert earphones (Akhoun et al., 2008), at a comfortable intensity of 50 dB SL (sensation level), with a rate of 1.1 per second, using a MATLAB (The MathWorks, USA) script to randomize stimulus polarities (1,500 presentations of each polarity). Participants were asked to remain awake during the recordings and count the number of times a green diode in front of them turned red (5% of the trials). The AEP recordings lasted 45 min with two breaks at 15 and 30 min.

Electrophysiological Recordings

Electrophysiological activity in response to the /ba/ syllable was recorded simultaneously by two EEG systems. A custom-made electronic device ensured that both systems were triggered simultaneously. The clinical Centor USB system (Racia-Alvar, France) recorded EEG signals from two derivations (sampling rate: 13.3 kHz; 70 Hz high-pass filter), both using the vertex (Cz) as the active electrode and one of the mastoids as the reference electrode (ground on the nasion), during 300-ms-long epochs triggered by the beginning of the sound file. The research BrainAmp system (Brain Products, Germany) recorded the EEG signal continuously (sampling rate: 5 kHz; 0.1 Hz high-pass filter) from 31 electrodes positioned in a cap following the 10–20 system, and one electrooculogram electrode at the outer canthus of the left eye—all 32 electrodes referenced to the nose (ground on the right cheek). Both systems used passive Ag/AgCl electrodes (impedance below 5 k Ω). Vertex and mastoid electrodes for both systems were positioned less than 1 cm away from each other.

Comparison of Speech ABRs Obtained With the Two EEG Systems

Data were analyzed using ELAN (Aguera, Jerbi, Caclin, & Bertrand, 2011) and MATLAB. For each participant, raw data from both EEG systems were first filtered with notch filters to remove the 50 Hz power line radiation artifact and its odd harmonics, then with an 80–1000 Hz Butterworth band-pass filter (fifth order) to isolate subcortical activity (additionally, to obtain cortical AEP from the BrainAmp data, a 3–80 Hz Butterworth band-pass filter, third order, was used after notch filters). Data from the BrainAmp system was epoched in 300-ms-long trials (for compatibility with the Centor data). For each system, trials that exceeded 60 μ V of amplitude dynamics at any electrode were rejected, together with trials during which the green diode turned red. Rejections were then combined between the two EEG systems to keep only trials acceptable on both systems. To compare the speech ABRs recorded with the two EEG systems, the signal at Cz was offline referenced for each system to the average of the two mastoids' signal (vertical montage). The signals from the two systems were compared in two ways: firstly, by calculating for each subject the cross-correlation between FFRs of the two signals (145 to 245 ms) and, secondly, by comparing the signal-to-noise ratios (SNR) for the two systems with a paired value *t* test. As the two EEG systems had different sampling rates, to compute the cross-correlation, the BrainAmp speech ABR was interpolated in the time base of the Centor speech ABR. We used the following formula to calculate SNR:

$10 * \log_{10} \left(\frac{EXY - EY}{EY} \right)$; with *EY* being the energy in the noise window (mean square [MS] of the 0–30 ms speech ABR portion)

and *EXY* being the energy of the signal + noise window (MS of the 150–180 ms speech ABR portion). SNR is expressed in decibels (dB).

Denosing Source Separation and Topography

We used denosing source separation (DSS) to improve SNRs of BrainAmp speech ABRs and, most importantly, to extract topographical information from speech ABR without having to choose any particular latency. DSS is a method that takes advantage of multichannel electrophysiological recordings by designing spatial filters to decompose the signal into as many components as electrodes (here, 31 for the BrainAmp data) using a criterion of stimulus-evoked reproducibility (de Cheveigné & Simon, 2008; Särelä & Valpola, 2005). Here, DSS was run on BrainAmp data on the same trials as those averaged to get speech ABR with a vertical montage, but on the 31-channel data (notch and band-pass filtered data files, original nose reference). We calculated the SNR for the best DSS component (representing the most reproducible part of the ERP) of each participant, and the spatial filter for this component corresponds to the scalp topography maximizing the evoked response. Finally, we ran DSS on subsets of the 500 and 1,000 first trials, and calculated SNR of resulting speech ABRs (along with SNRs of vertical montage averaging of the 500 and 1,000 first trials), to assess the potential role of this method in reducing the recording time.

Results

Speech ABRs evoked by the /ba/ stimulus (Figure 1a) are shown in Figures 1b,c for two participants. Visual inspection of the curves suggests similar noise levels and time courses of signal for both systems. Expected components of speech ABR are retrieved with both systems, with an onset response to /b/ release at 150 ms, and a FFR to vowel periodicity between 150 and 225 ms. Across all participants, maxima of the cross-correlation of speech ABRs obtained with the two systems were found at a constant lag of around 0.75 ms, which may be the reflection of online filter differences between the two EEG systems. We found an average r^2 of .39 (Pearson correlation coefficient, see Table 1; the correlation for each participant was significant, $p < .0001$), showing high resemblance between the two signals as they share almost 40% of their variance. Moreover, comparison of SNRs between speech ABRs obtained with clinical and research EEG systems revealed no significant difference (paired *t* test: $t(13) = 1.448$, $p = .17$). Altogether, visual inspection of the temporal courses of ERPs, cross-correlation, and SNR measurements for the two EEG systems support the view that it is possible to record speech ABR with a classic multichannel research EEG system.

As a consequence, it is possible to benefit from all the information provided by such a system: using information on the 31 channels for advanced processing or analysis methods, but also getting cortical AEP and their topography (Figure 2d), and for the first time topographical information on speech ABR. Figure 2b,c, respectively, shows grand averages of speech ABRs ($N = 14$) observed at Cz rereferenced to the average of the mastoids (classical vertical montage) and speech ABR obtained with DSS (allowing us to obtain an optimal montage corresponding to speech ABR topography), both evoked by the /ba/ stimulus shown in Figure 2a. Speech ABR topography (Figure 2c) remarkably emphasizes the usefulness of the classical vertical montage (vertex–mastoids), but goes further using information on the 28 other electrodes.

Table 1. Comparison of Speech ABRs Obtained with the Two EEG Systems, and Benefits of Using DSS

	Number of accepted trials	r^2	SNR (dB)		
			Centor USB	BrainAmp	BrainAmp + DSS
S1	1,592	.41	12.2	12.8	19.4
S2	1,777	.56	5.4	8.0	12.2
S3	1,714	.61	5.8	1.2	8.6
S4	1,593	.33	4.2	-11.1	-3.8
S5	1,650	.21	3.3	1.3	-2.4
S6	1,648	.02	1.5	8.0	12.3
S7	1,335	.33	-5.7	-2.7	-0.6
S8	1,710	.48	8.5	5.7	8.0
S9	1,719	.44	9.2	3.4	6.9
S10	1,611	.27	5.1	8.1	9.3
S11	1,775	.50	11.2	9.6	10.8
S12	1,264	.65	10.2	5.8	6.7
S13	2,483	.52	17.1	18.1	19.8
S14	1,531	.19	8.6	-4.5	5.0
Mean	1,672	.39	6.9	4.5	8.0

Note. For each participant (S1 to S14), r^2 is reported at a 0.75-ms lag (Centor ahead of BrainAmp data), to take into account filter differences between the two systems (see text for details). SNR is reported for the classical vertical montage (Centor and BrainAmp columns) and for the best DSS component (BrainAmp data only). SNR = signal-to-noise ratio; DSS = denoising source separation.

The multichannel denoising method tested here, DSS, obviously increases SNR as confirmed in Table 1 (paired t test: $t(13) = 3.788$, $p = .002$). Furthermore, as shown in Figure 2e, the SNR obtained using a vertical montage by averaging all acceptable trials (mean 1,672 across participants) from the research EEG system is reached and even exceeded in only 1,000 trials using DSS. Using DSS on the 500 first trials led to speech ABRs with an acceptable SNR of almost 3 dB (Figure 2e), as is the case for the clinical EEG system.

Discussion

The present article reports a new methodology for multichannel recording of speech ABR with a critical comparison of the same signal recorded with two different EEG systems; to the best of our knowledge, this is the first control of the effect of lowering sampling rate on speech ABR's quality. We show here that it is possible to record high-quality speech ABR using a research EEG system with a relatively low sampling rate (here, 5 kHz). The speech ABRs thus recorded were indistinguishable in terms of SNR from speech ABRs acquired with a clinical EEG system as is most often done in speech ABR studies. Note that SNRs in this study were comparable and even superior on average to SNRs measured in other studies, which range typically between 2.5 and 3 dB (Skoe & Kraus, 2010),

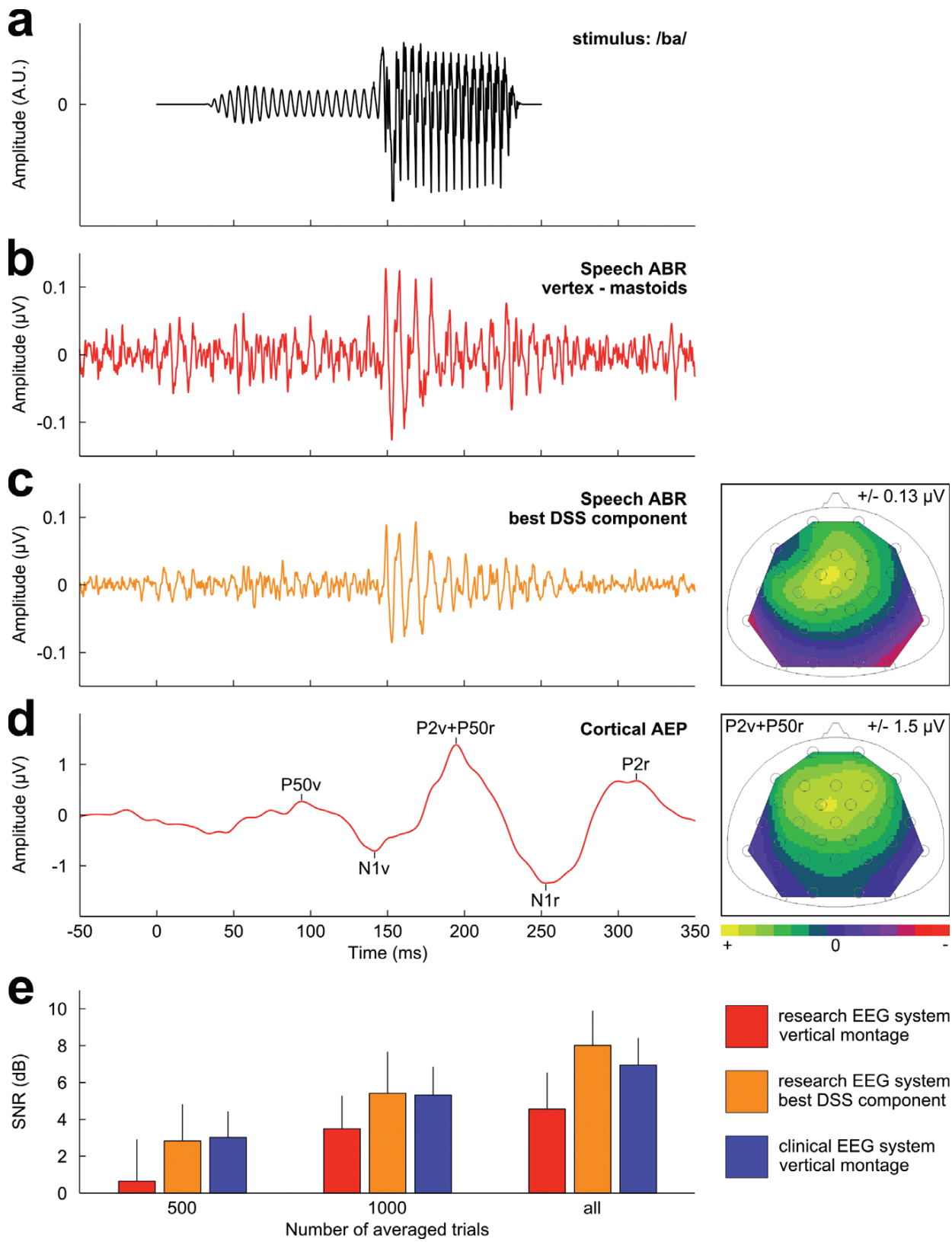
whatever the EEG system or data processing method used, thus confirming the validity of our results.

We retrieved here a similar mean correlation coefficient between the speech ABRs recorded with a clinical and a research EEG system as in a within-session test-retest reliability study of speech ABR (Hornickel, Knowles, & Kraus, 2012; note, however, that 8- to 13-year-old children were tested in this previous report). In the test-retest study, the correlation coefficient between two speech ABRs from two separate runs of 3,000 trials in the same session was on average 0.71 (Spearman's ρ), with a standard deviation (SD) of 0.3. Here, mean Spearman's ρ across our 14 participants is 0.59 ($SD = 0.17$). The small difference could be explained by the different positions (1 cm) between identical electrodes of both EEG systems, and by the slightly lower number of trials in the present study.

A variant of alternating polarities, consisting of subtracting speech ABR to both polarities instead of summing them, put the emphasis on the high frequency components of the signal, without minimizing possible stimulus artifact or cochlear microphonics contribution to the response. Using this variant (Aiken & Picton, 2008), there might be an advantage to record speech ABR with a very high sampling rate EEG system, but as the large majority of speech ABR studies uses the classical alternating polarities approach (summation), the interest of our methodology is still valid.

Thus, the present methodology enhances classical speech ABR recording, by allowing the simultaneous collection of cortical AEP and speech ABR topographies. In other words, it provides a wide panorama of electrophysiological activities evoked by a speech stimulus within a single recording session. Moreover, topographies of speech ABR have never been described, although they may represent an interesting electrophysiological marker in addition to latency and amplitude measurements, and allow the experimenter to gain insight into the poorly known generator(s) of this response. Note that a few reports of click ABR topography can be found in the literature (e.g., Norrix & Glatke, 1996). Using this methodology, the only drawback appears to be the time of electrodes setup (30 min instead of 10, which could be reduced in the future thanks to the use of dry electrodes and/or wireless EEG systems) and the increased data size (a less problematic issue nowadays). Simultaneous recordings of speech ABR and cortical AEP have been reported in only one study so far, but without any topographic EEG analysis (Musacchia, Strait, & Kraus, 2008). We agree with its authors on the fact that these two types of AEPs have different optimal stimulation parameters (number of presented trials and interstimulus interval). Although this might result in a drastically increased recording time to get both responses simultaneously, the issue can be bypassed by acquiring both markers in separate runs with respective optimal parameters. Furthermore, as shown in Figure 2e, multichannel recording of speech ABR allows using advanced processing methods like DSS, which allows for the

Figure 2. Multiple benefits of topographic recordings of AEP to speech (stimulus in a). Data shown in b, c, and d are grand averages of the fourteen participants' ERPs. With the classical vertical montage (b), we observe both onset response and FFR as described for individual participants in Figure 1. With the optimal montage obtained using DSS (c), signal-to-noise ratio is improved, and we get for the first time topographic information on subcortical ERP to speech (right panel): a frontocentral positive spot with polarity inversion at mastoids with a nose reference. Concomitant cortical AEP (d) presents two P50-N1-P2 wave sequences in response to voicing (v) and /b/ release (r), with P2 to voicing (P2v) superimposed with P50 to /b/ release (P50r). Topography at 200 ms (right panel) is typical of cortical AEPs. Note that amplitude for cortical AEPs is ten times higher than for speech ABR. Finally in (e), we show the SNR for the classical vertical montage for the research and clinical EEG systems or for the best DSS components when averaging the first 500, the first 1,000, or all trials (1,672 on average; the bars represent the mean SNR across participants $\pm SEM$).



reaching of high SNRs in only half the averaged trials required for classic processing. This is in agreement with a recent study using frequency-domain principal component analysis on multichannel recording of FFR to click trains, thus reducing by a factor of 3 the number of trials needed (Bharadwaj & Shinn-Cunningham, 2014).

Further studies are needed for assessing speech ABR generators using the present methodology. In the future, combining the use of

speech ABR topographies and concomitant cortical AEPs should allow a better characterization of speech processing in populations with various hearing abilities. In any case, by proving the feasibility of recording high-quality speech ABR with standard EEG systems, and considering the widespread use of research EEG systems in neuroscience laboratories as well as the numerous fields of application of speech ABR, the present findings should amplify the development of speech ABR studies.

References

- Aguera, P.-E., Jerbi, K., Caclin, A., & Bertrand, O. (2011). ELAN: A software package for analysis and visualization of MEG, EEG, and LFP signals. *Computational Intelligence and Neuroscience*, 2011, 158970. doi: 10.1155/2011/158970
- Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology & Neurotology*, 11, 213–232. doi: 10.1159/000092589
- Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, 245, 35–47. doi: 10.1016/j.heares.2008.08.004
- Akhoun, I., Moulin, A., Jeanvoine, A., Ménard, M., Buret, F., Vollaïre, C., . . . Thai-Van, H. (2008). Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: An experimental parametric study. *Journal of Neuroscience Methods*, 175, 196–205. doi: 10.1016/j.jneumeth.2008.07.026
- Anderson, S., & Kraus, N. (2013). The potential role of the cABR in assessment and management of hearing impairment. *International Journal of Otolaryngology*, 2013, 604729. doi: 10.1155/2013/604729
- Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady state responses using multichannel recordings. *Clinical Neurophysiology*, 125, 1878–1888. doi: 10.1016/j.clinph.2014.01.011
- Dajani, H. R., Heffernan, B. P., & Giguere, C. (2013). Improving hearing aid fitting using the speech-evoked auditory brainstem response. *IEEE Transactions on Biomedical Engineering*, 2013, 2812–2815. doi: 10.1109/EMBC.2013.6610125
- Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52, 1614–1618. doi: 10.1109/TBME.2005.851499
- de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, 171, 331–339. doi: 10.1016/j.jneumeth.2008.03.015
- Galbraith, G. C., Threadgill, M. R., Hemsley, J., Salour, K., Songdej, N., Ton, J., & Cheung, L. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neuroscience Letters*, 292, 123–127. doi: 10.1016/S0304-3940(00)01436-1
- Hornickel, J., Knowles, E., & Kraus, N. (2012). Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hearing Research*, 284, 52–58. doi: 10.1016/j.heares.2011.12.005
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, 241, 34–42. doi: 10.1016/j.heares.2008.04.013
- Norrix, L. W., & Glatke, T. J. (1996). Multichannel waveforms and topographic mapping of the auditory brainstem response under common stimulus and recording conditions. *Journal of Communication Disorders*, 29, 157–182. doi: 10.1016/0021-9924(95)00019-4
- Parbery-Clark, A., Strait, D. L., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *Journal of Neuroscience*, 33, 16741–16747. doi: 10.1523/JNEUROSCI.5700-12.2013
- Prévost, F., Laroche, M., Marcoux, A. M., & Dajani, H. R. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology*, 124, 52–60. doi: 10.1016/j.clinph.2012.05.009
- Richard, C., Jeanvoine, A., Veuillet, E., Moulin, A., & Thai-Van, H. (2010). [Exploration of the auditory system in humans: From click to speech auditory brainstem responses]. *Neurophysiologie Clinique*, 40, 267–279. doi: 10.1016/j.neucli.2010.08.002
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119, 1720–1731. doi: 10.1016/j.clinph.2008.01.108
- Särelä, J., & Valpola, H. (2005). Denoising source separation. *Journal of Machine Learning Research*, 6, 233–272.
- Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing*, 31, 302–324. doi: 10.1097/AUD.0b013e3181c8b272

(RECEIVED March 11, 2014; ACCEPTED September 7, 2014)

4.2. Composante corticale de la FFR

L'article intitulé [Une composante corticale de la Frequency-Following Response évoquée par la parole : preuves en SEEG] (*A cortical component in the Frequency-Following Response to Speech: evidence from SEEG*) est actuellement en voie de soumission dans le journal à comité de lecture *Proceedings of the National Academy of Sciences*.

A cortical component in the Frequency-Following Response to Speech: evidence from intracranial EEG.

Bellier L, Bidet-Caulet A, Bertrand O, Catenoix H, Thai-Van H & Caclin A

Keywords

Electrophysiology; auditory evoked potentials; phase-locked activity; auditory cortex; inferior colliculus; intracranial EEG; scalp EEG; neural generator; scalp current density

Abstract

Neural generators of the Frequency-Following Response (FFR) to speech sounds, a component of the Speech Auditory Brainstem Response (Speech ABR), have been assumed to lie at subcortical level. However, FFR to fundamental frequencies (F0) up to 114 Hz have recently been observed within human primary auditory cortex (A1). In a first part of the present study, we assessed the existence of a cortical FFR to speech for higher F0, using intracranial SEEG recordings in normal-hearing patients with epilepsy. Cortical FFR was indeed present within bilateral A1 for speech F0 up to 200 Hz, in seven out of nine patients implanted with electrodes in superior temporal auditory cortices. Cortical FFR recorded intracranially occurred simultaneously to scalp-recorded Speech ABRs evoked by the same speech *stimuli*. In a second part, we investigate the possible contribution of cortical FFR in generating the scalp-recorded response using multichannel scalp recordings. Scalp Current Density (SCD) topographies of the scalp-recorded FFR exhibited bilateral temporal dipoles, similar to those observed for the classical P50 wave. Together, our results provide the first evidence for a cortical generator of Speech ABR's FFR, thus changing the way we think of this electrophysiological marker of speech encoding.

Significance Statement

For the last 20 years, Speech Auditory Brainstem Response has been considered to be generated at a subcortical level. In our present study, SEEG intracranial recordings revealed the existence of Frequency-Following Response, one of the Speech ABR's components, within bilateral primary auditory cortex and for fundamental frequencies up to 200 Hz. Furthermore, scalp-EEG topographies suggested a participation of this cortical FFR in generating the scalp-recorded FFR. Together, our findings unveil a composite nature of Speech ABR, with both subcortical and cortical generators, and lead to a reinterpretation of dozens of Speech ABR studies.

Introduction

Speech Auditory Brainstem Response (Speech ABR; also called complex ABR) is known as a scalp-recorded electrophysiological marker of speech encoding at the brainstem level (Skoe & Kraus, 2010). Over the last two decades, it has increasingly been studied in relation to many topics, such as music (Wong *et al.*, 2007; Musacchia *et al.*, 2007; Bidelman & Krishnan, 2009; Kraus & Chandrasekaran, 2010; Bidelman & Krishnan, 2011; Strait & Kraus, 2014), language (Krishnan *et al.*, 2005; Krishnan *et al.*, 2010; Krizman *et al.*, 2012), speech-in-noise (Hornickel *et al.*, 2009; Anderson & Kraus, 2010; Anderson *et al.*, 2010; Song *et al.*, 2011), aging (Anderson *et al.*, 2012; Clinard & Tremblay, 2013; Bidelman *et al.*, 2014), autism (Russo *et al.*, 2008; Russo *et al.*, 2009) and learning disorders (Cunningham *et al.*, 2001; King *et al.*, 2002; Wible *et al.*, 2004), among others. Yet, despite this growing interest for Speech ABR, only two studies investigated its neural generators, both reaching the same conclusion: Speech ABR is a subcortical Auditory-Evoked Potential (AEP; Chandrasekaran & Kraus, 2010; Bidelman, 2015)—hence brainstem in its acronym. Differences between Speech ABR of distinct populations (e.g., musicians and non-musicians), or evolution of Speech ABR parallel to improvement of auditory-related skills (e.g., auditory training), are thus currently interpreted as consequences of neuronal plasticity at the brainstem level.

Speech ABR is a composite of two types of neural activity, considered equivalents of two classical AEPs: Click ABR and Frequency-Following Response (FFR; Skoe & Kraus, 2010). Transient components, namely onset and offset response, happen approximately six milliseconds after an abrupt change of energy level in *stimulus* acoustics. They are compared to peak V of Click ABR (Jewett *et al.*, 1970; Jewett & Williston, 1971), known to be generated by the rostral part of the lateral lemniscus, just before reaching the inferior colliculus (IC; Scherg & Von Cramon, 1985; Møller *et al.*, 1994; Møller, 2007). Sustained component, namely FFR—as the eponymous classical AEP, is a phase-locked activity which reflects several spectral features of speech sounds (fundamental frequency, harmonics, and formants; Moushegian *et al.*, 1973; Krishnan, 1999; Plyler & Ananthanaryan, 2001; Krishnan, 2002; Johnson *et al.*, 2008), and occurs approximately 15 ms after the beginning of a periodic part in the speech *stimulus* (Akhoun *et al.*, 2008). Neural generators of FFR to speech have almost never been studied, but are rather assumed to be the same as those of the classical 6-ms-latency FFR to tone bursts, which have been found at different levels of the auditory pathways (cochlea's hair cells with their cochlear microphonics, cochlear nucleus and IC; Worden & Marsh, 1968; Gerken *et al.*, 1975; Sohmer & Pratt, 1977; Sohmer *et al.*, 1977; Stillman *et al.*, 1978). As these generators have distinct locations and orientations, they can be separately highlighted by different setup configurations and stimulation strategies. The most common recording method in Speech ABR studies emphasizes phase-locking activity to fundamental frequency (F0) of the speech sound, by combining a vertical montage (vertex referenced to linked mastoids) with alternating polarity (average of evoked-responses to both positive and negative versions of the *stimulus*). Resulting FFR is then called envelope FFR (Aiken & Picton, 2008), and is widely accepted to be generated in the upper brainstem (namely IC; Wong *et al.*, 2007; Kraus & Chandrasekaran, 2010; Chandrasekaran & Kraus, 2010; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011a,b; Anderson *et al.*, 2012; Krizman *et al.*, 2012; Bellier *et al.*, 2015a; Bidelman, 2015).

Yet, longer latency of Speech ABR's FFR (approx. 15 ms) compared to tone-evoked FFR (approx. 6 ms) is compatible with a cortical generator, as the first cortical auditory responses reach the primary auditory cortex (A1) as soon as 13 to 19 ms after *stimulus* onset (Liégeois-Chauvel *et al.*, 2001; Yvert *et al.*, 2002; Lütkenhöner *et al.*, 2002; Parkkonen *et al.*, 2009). Moreover, in another field of neuroscience dedicated to studying how the human primary auditory cortex (A1) works, phase-locking activity has been measured in response to speech for F0 up to 114 Hz, within human A1, in patients with epilepsy undergoing neurosurgical procedure (Steinschneider *et al.*, 2013). Thus, A1 is able to exhibit a FFR activity, at least up to 114 Hz.

In the present study, we ask two questions: does FFR to speech F0 exists within the human A1, beyond the currently known upper limit of 114 Hz? If it does, could cortical FFR participate in generating the scalp-recorded Speech ABR's FFR?

We addressed the first question by recording electrophysiological activity evoked by two /ba/ speech syllables with F0 of 170 and 200 Hz (respectively *ba170* and *ba200*), in nine patients with intractable epilepsy undergoing a Stereo-Electroencephalography (SEEG) clinical procedure to localize their epileptic foci before surgical removal. Classical scalp-recorded Speech-ABR protocol was adapted to fit clinical requirements, and SEEG data was processed and analyzed in the same way as scalp-recorded Speech ABR, except regarding contact referencing. Indeed, we used a bipolar montage, where each contact of an electrode serve as a reference to its nearest neighbor. This minimizes contribution of distant sources of neural activity (or non-physiological activity such as *stimulus* artifact), and improves the spatial resolution of SEEG to 5 mm around bipolarized virtual contacts (Lachaux *et al.*, 2003; Lachaux *et al.*, 2012).

To address the second question, we re-analyzed data of two of our previous scalp-recorded Speech ABR studies, using *ba170* and *ba200* as *stimuli* respectively (Bellier *et al.*, 2013; Bellier *et al.*, 2015a). Scalp topography of Speech ABR's FFR was computed for both studies using Scalp Current Density (SCD; second derivative of the EEG potential, also known as surface Laplacian). SCD being fairly insensitive to deep neural generators such as IC in the upper brainstem, observing activity on SCD would suggest cortical generators (Perrin *et al.*, 1987; Pernier *et al.*, 1988). As a positive control for this method, SCD of the speech-evoked FFR was compared to SCD of the classical cortical P50 (Liégeois-Chauvel *et al.*, 1994; Yvert *et al.*, 2001).

Results

SEEG results. A first look into Intracranial AEPs. Using bipolar referencing and alternating polarity, intracranial recordings allowed for observing AEPs in response to speech sounds, at numerous contacts and within different frequency bands. Unfiltered raw data were dominated by low-frequency intracranial AEPs, present on a far larger amount of contacts than AEPs of higher frequencies (Table 1); yet, most of these contacts exhibited slow drifts unspecific of syllables' acoustics, while the most complex shapes were found within auditory cortices. High-pass filtering unmasked activities of lesser amplitude, more reactive to tiny sound variations.

Table 1. SEEG contacts with significant activity

Patient	Total nb. of contacts	RAW		HP80		BP160-180		BP190-210	
		<i>ba170</i>	<i>ba200</i>	<i>ba170</i>	<i>ba200</i>	<i>ba170</i>	<i>ba200</i>	<i>ba170</i>	<i>ba200</i>
○ P1	152	64	81	12	13	9	0	0	4
□ P2	105	38	40	8	8	8	5	0	6
◇ P3	155	19	33	0	0	0	0	0	0
▽ P4	118	55	55	11	8	9	3	4	6
△ P5	120	26	22	5	4	4	0	1	1
◁ P6	181	85	40	6	8	1	0	0	1
▷ P7	92	34	23	7	7	7	6	0	6
☆ P8	135	89	64	6	8	4	0	0	3
☆ P9	115	34	32	0	0	0	0	0	0
Total	1173	444	390	55	56	42	14	5	27

Number of contacts showing significant activity for each patient, *stimulus* and filtering. Only contacts with more than one significant sample within a temporal window of interest (0-to-250 ms for *ba170*; 0-to-300 ms for *ba200*) were considered as significant. RAW corresponds to unfiltered data.

Existence of a cortical FFR similar to the scalp-recorded FFR. Among high-frequency AEPs, we found an activity specifically phase-locked to F0 of both *ba170* and *ba200* syllables (Fig. 1 & Table 1), thus corresponding to the definition of the Frequency-Following Response: FFR does seem to exist at a cortical level, for F0 at least up to 200 Hz.

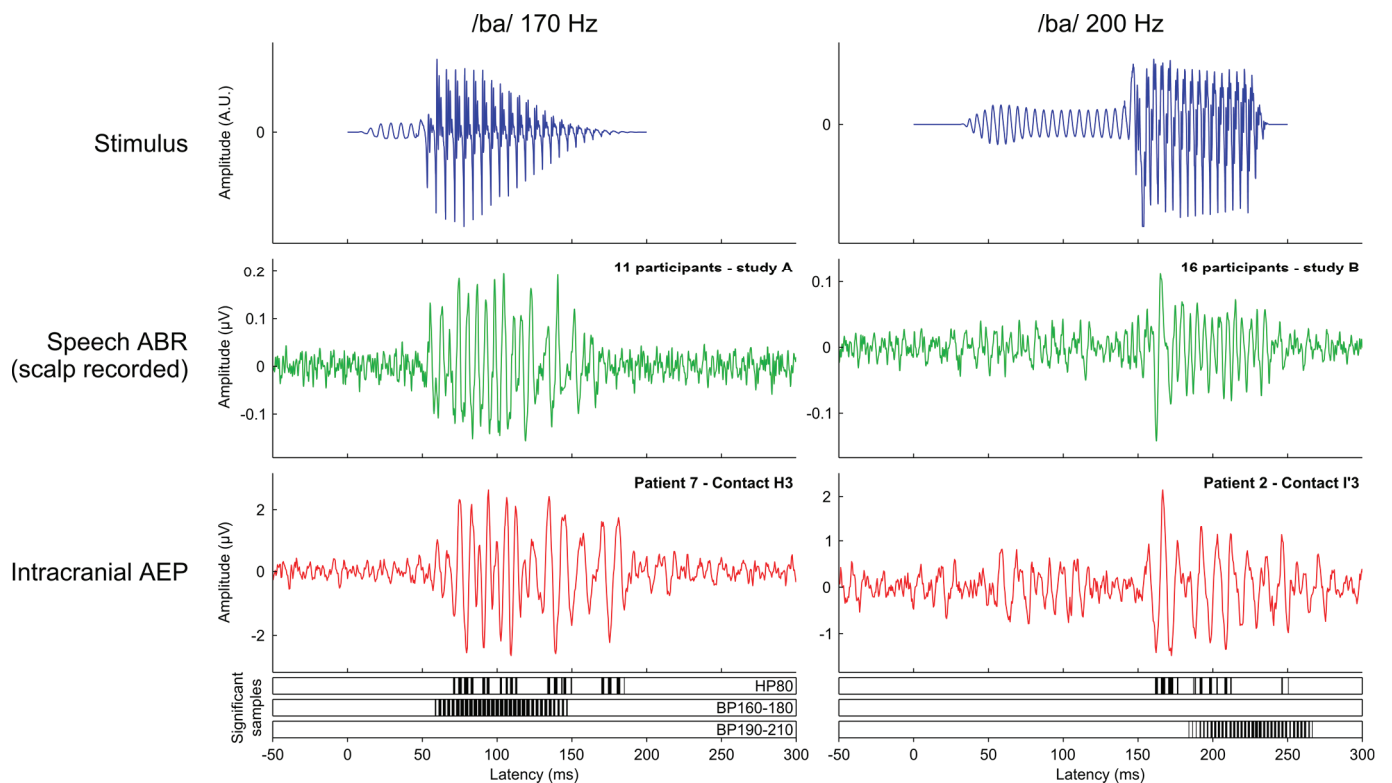


Fig. 1. Comparison between scalp-recorded Speech ABR and intracranial speech-evoked AEP exhibiting the cortical FFR. Data evoked by both *ba170* and *ba200* stimuli are presented on left and right columns, respectively. Below intracranial AEP, samples significantly emerging from baseline are shown for different filtering ($p < 0.05$, FDR-corrected; HP80 = 80-Hz high-pass filter; BP160-180 = 160-to-180-Hz band-pass filter; BP190-210 = 190-to-210-Hz band-pass filter).

Waveforms of these so-called cortical FFRs were visually similar to their related scalp-recorded FFRs, and occurred simultaneously. Cortical FFR was found in seven out of the nine recorded patients (Table 1), for a total of 42 contacts in response to *ba170* and 27 in response to *ba200*, respectively over 55 and 56 contacts showing phase-locking activity above 80 Hz in response to speech sounds. Moreover, some contacts also showed F0-unspecific high-frequency evoked activity. For example, in response to *ba170*, five contacts out of the 42 manifesting a cortical FFR—with significant evoked activity in the 160-to-180-Hz frequency band, also shown F0-unspecific evoked activity in the 190-to-210-Hz frequency band. Note that to our knowledge, this is the first time that a cortical FFR is observed for F0 as high as 170 or even 200 Hz, the previous limit being 114 Hz (Steinschneider *et al.*, 2013).

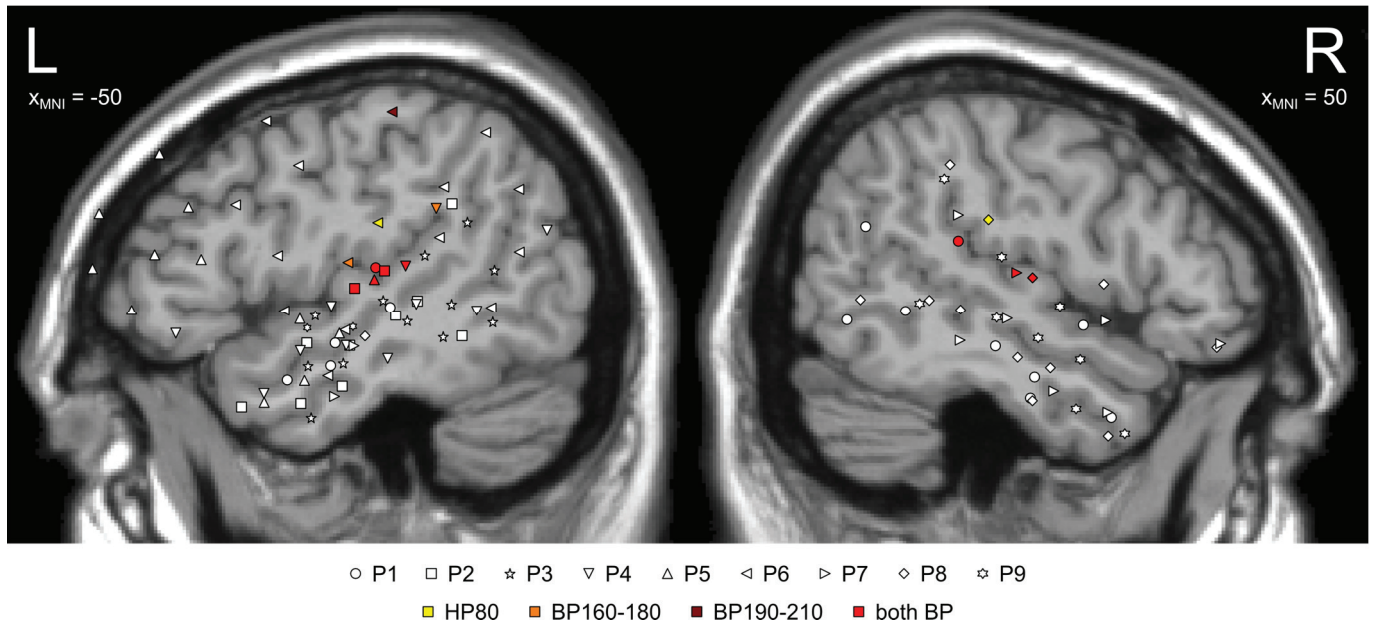


Fig. 2. Location of all electrodes for all patients. Electrodes are plotted on sagittal slice of a T1 MNI template. Patient identities are represented with the same shapes as in Table 1. Colors indicate the presence of at least one contact showing significant activity in both frequency bands (BP160-180 and BP190-210), in only one, or at frequencies superior to 80 Hz, in order of decreasing priority to define contact's color.

Location of contacts showing a cortical FFR. Location of electrodes with contacts showing a FFR to both *stimuli* was restricted to medial and posterior parts of bilateral Superior Temporal Gyrus (STG; Fig. 2, contacts in red). Looking more closely at an electrode penetration in patient P1 (Fig. 3), we found that location of contacts showing a cortical FFR to both *stimuli* could even be specified to posteromedial portion of Heschl's gyrus (HG), where A1 lies (Liégeois-Chauvel *et al.*, 1991). Interestingly, some surrounding contacts also showed a cortical FFR to *ba170*, but not to *ba200*, and F0-unspecific high-frequency activity was present on even more lateral contacts.

Restricted location of cortical FFR to posteromedial A1 was also observed in other patients (Fig. 4, in red). Moreover, electrode location on individual anatomy helped to better understand inter-individual variability in the presence or not of cortical FFR in the recordings: in P3, H' electrode was not deep enough to reach A1; in P6, H' electrode was too anterior within the supra-temporal plane and only reached lateral HG. However, absence of cortical FFR on the

H electrode of patient P9 could not be explained by electrode location, as there were contacts well through posteromedial A1; instead, this might be due to an abnormal cortical architecture in this epileptic patient, and/or to poor pitch-tracking, as sometimes observed in scalp-recorded Speech ABR's FFR.

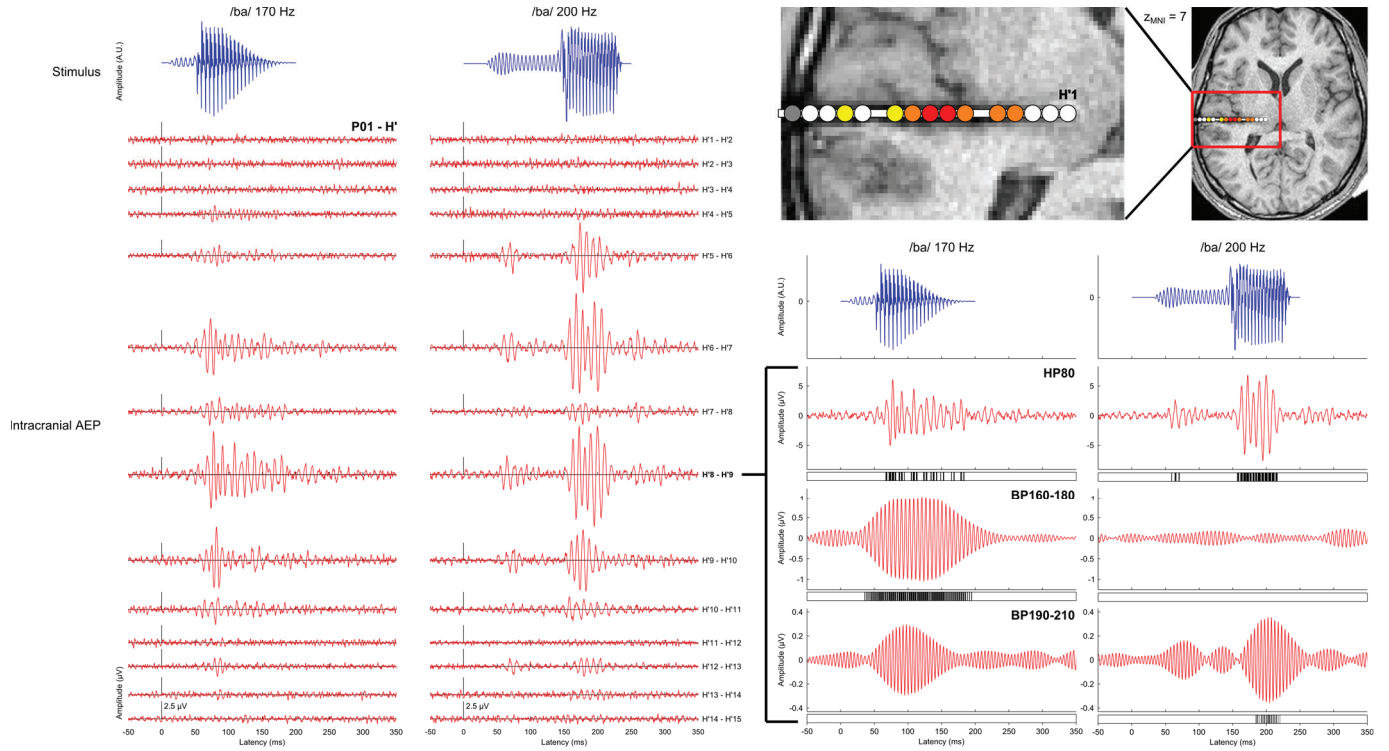


Fig. 3. Intracranial AEPs along H' electrode of P1. Same shapes and colors as in Fig. 2. Band-pass-filtered versions of cortical FFR are shown for bipolar contact H'8-H'9, with significant samples indicated as in Fig. 1.

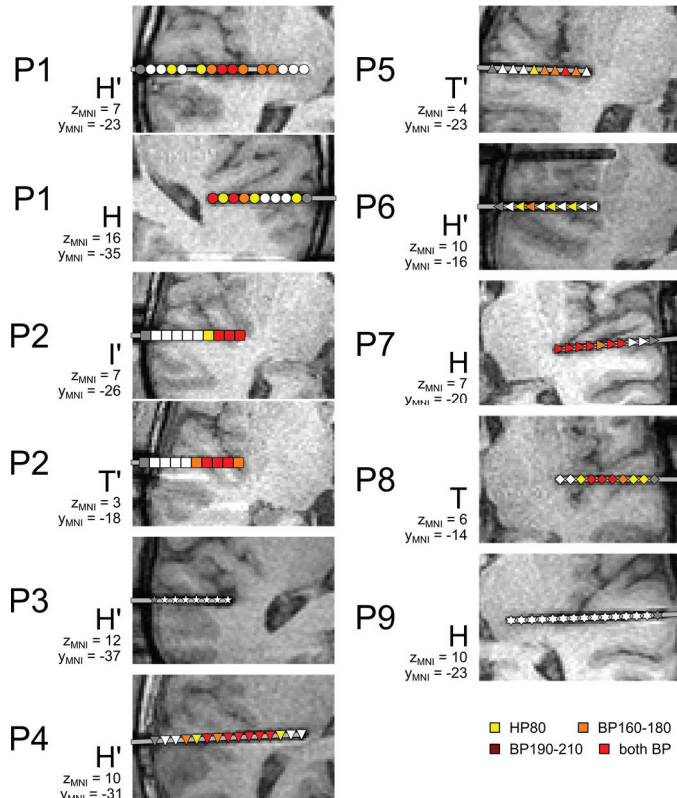


Fig. 4. Electrode implantations on individual anatomy for all patients. Same shapes and colors as in Fig. 2. Electrodes are plotted on transversal slices, and MNI coordinates z and y were indicated to help identifying electrodes on Fig. 2.

The present SEEG results thus demonstrate the existence of a cortical FFR, at least within bilateral posteromedial HG (A1). Cortical FFR was observed in seven out of nine patients, for F0 up to 200 Hz. Two points should however be highlighted here. First, existence of FFR activity in A1 does not imply the non-existence of FFR activity in brainstem nuclei (mainly IC). Second, cortical FFR does not necessarily participate in generating the scalp-recorded FFR, as dipole antagonisms could keep it from diffusing to the scalp, or the resulting scalp response might be of a very small amplitude compared to other generators.

Scalp-recorded Speech ABR results. To assess the potential contribution of A1 in generating the Speech ABR's FFR, we compared SCD of scalp-recorded FFR to SCD of scalp-recorded Cortical AEP, using the data of two of our previous topographic EEG studies using *ba170* and *ba200* as *stimuli*. We used the P50 wave topography as a within-study reference for auditory cortex activity. As expected, and in both studies, SCD topography of cortical P50 exhibited bilateral temporal dipoles, reflecting evoked activity within A1 and other auditory areas. In both studies, we found similar dipoles in SCDs of the scalp-recorded FFR to speech, thus suggesting bilateral cortical generators of this thought-to-be subcortical response (Fig. 5).

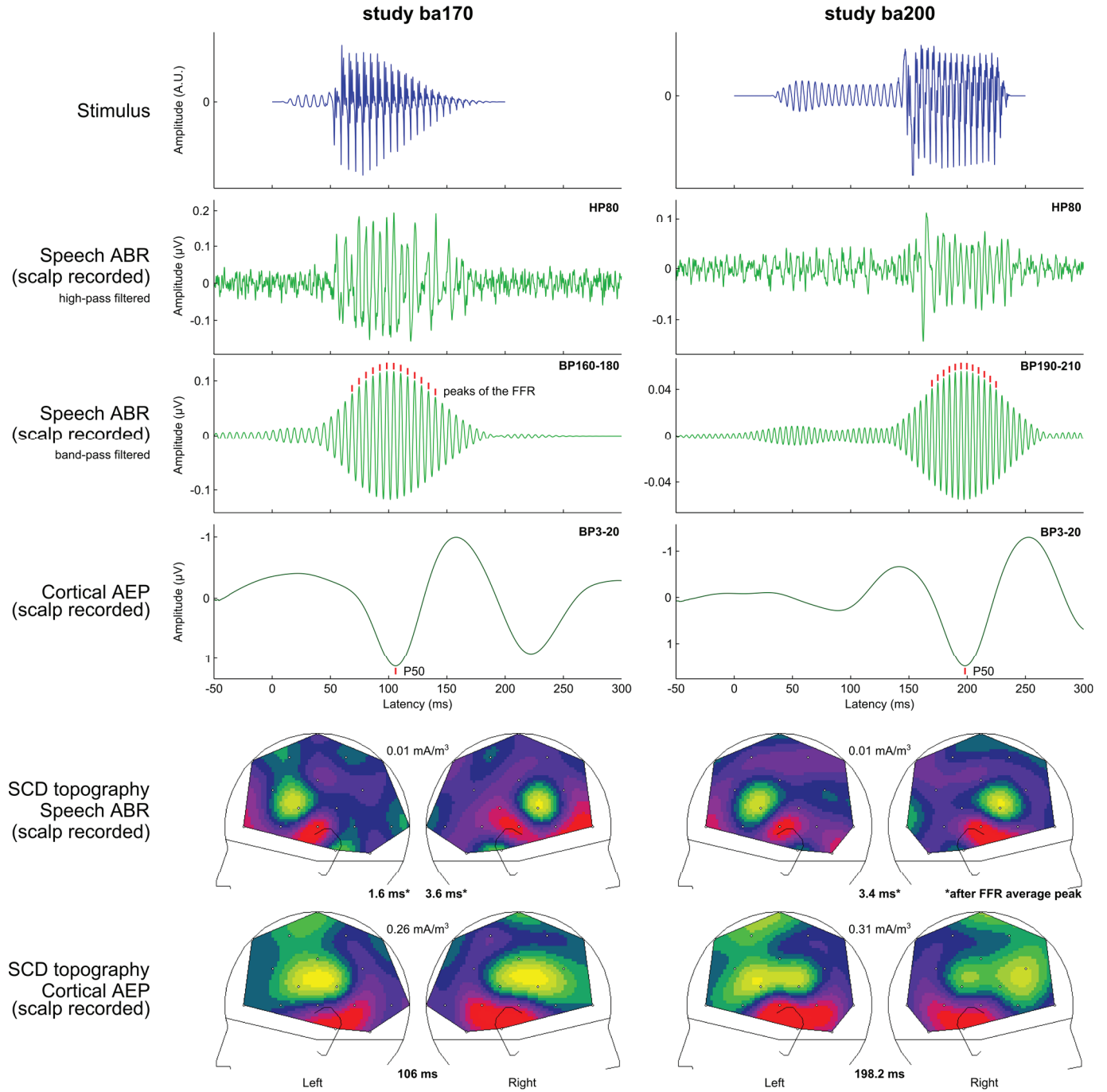


Fig. 5. Comparison between SCD topographies of the scalp-recorded FFR and P50. Red marks above band-pass-filtered Speech ABRs indicate the FFR peaks composing the average.

Interestingly, these temporal dipoles reached their maxima at different latencies than the response peak at the vertex electrode (occurring at 0 ms), which could reflect the subcortical contribution of upper brainstem (IC) in Speech ABR's FFR generation. Together, these results are compatible with the assumption of a participation of the SEEG-observed cortical FFR in generating the scalp-recorded Speech ABR's FFR. To our knowledge, this constitute the first proof of a cortical component of speech-evoked FFR to F0, which could pave the way to new interpretations of Speech ABR studies.

Discussion

SEEG results reveal for the first time the existence of speech-evoked FFR to F0 up to 200 Hz, in bilateral A1. Additionally, scalp-EEG studies suggest a participation of this cortical FFR in generating the scalp-recorded FFR component of Speech ABR. Thus, FFR of the Speech Auditory Brainstem Response may not be of subcortical nature only, but instead a composite of subcortical and cortical phase-locking activity. From this perspective, group or training effects on Speech ABR's envelope FFR, reported by many studies for the last 20 years, could not be interpreted anymore as a single manifestation of brainstem plasticity, as changes could rather (or also) come from the cortical component of FFR. Our results do not mean that there is no FFR component of subcortical nature, or that brainstem plasticity does not exist; our present study simply suggest that Speech ABR is not only generated in the brainstem, but also in the auditory cortex.

Existence of a physiological cortical FFR to F0 of *ba170* and *ba200* in bilateral A1 is demonstrated here thanks to SEEG recordings, which allow to exclude a contribution of activity from IC or other subcortical nuclei as a source of the recorded response, or a *stimulus* artifact. Firstly, bipolar referencing improves spatial resolution to about 5 mm around recording site, thus discarding potential contributions from IC or other subcortical nuclei. Secondly, cortical FFR to both *stimuli* was only observed within A1, inconsistently with distal contribution of IC or *stimulus* artifact hypotheses: we would instead observe FFR on contacts closer to IC or on a lot more contacts, respectively. Thirdly, alternating polarity minimizes the *stimulus* artifact (Campbell *et al.*, 2012), as average of positive and negative artifact is zero. At last, as shown in Figure 3, FFR activity was changing across contacts of a same electrode; *stimulus* artifact would instead result in almost identical waveforms.

Thus, this is the first time to our knowledge that FFR to speech is directly recorded from within human A1 for F0 upper than 114 Hz. This constitutes a solid proof that human A1 is able to phase-lock spectral components up to at least 200 Hz, while its phase-locking cutoff frequency was accepted to be around 100 Hz (Chandrasekaran & Kraus, 2010; Bidelman, 2015).

In addition to cortical FFR to F0 of the two /ba/ syllables observed in bilateral A1 only, three contacts also exhibited FFR to one of the two *stimuli* only, out of A1. In P4, 25-ms-long significant FFR activity was found for *ba170* only, on the G'7-G'8 bipolar contact located in the white matter of the parietal lobe near the left lateral ventricle ($xyz_{MNI} = [-32 -40 24]$). In P6, 49-

ms-long FFR to *ba200* only was found on Z'4-Z'5 contact located in between the left pre- and post-central gyri ($xyz_{MNI} = [-30 -46 54]$); also, and as exposed in the results, H'7-H'8 contact showed a 48-ms-long FFR to *ba170* only, and was located within anterolateral HG ($xyz_{MNI} = [-57 -16 10]$). The last one suggests that phase locking persists beyond A1, yet with a decreasing cutoff frequency when the information is transmitted further up in the auditory hierarchy. Small FFR-like responses in somatosensory and motor cortices, as observed in P6 could participate in representing some articulatory correlates of the speech *stimuli*. In any case, these AEPs had durations of significant activity long enough not to consider them as false positives; yet, their spatial isolation along with their sensitivity to one but not both speech syllables call for caution in their interpretation.

High-frequency evoked activity unspecific of *stimulus* F0 was also observed on several contacts, particularly in auditory cortices (see yellow marks in Figures 2-4), and might also be found in scalp-recorded activity when using an 80-Hz high-pass filter, suggesting that such a filter does not isolate subcortical from cortical activity as classically assumed in Speech ABR literature (Skoe & Kraus, 2010).

At the scalp level, FFR is known to be a composite of phase-locking activity to different spectral components of the *stimulus* (F0, harmonics and formants), generated at different steps of the auditory pathways (cochlea, auditory nerve, cochlear nucleus, IC, and now A1). It is of importance to notice that we focus here on FFR to speech F0. Human auditory system also uses phase-locking to encode harmonics (multiples of F0) and formants (modulations of spectral amplitude of F0 and harmonics, underlying phoneme perception; Plyler & Ananthanaryan, 2001), necessarily of higher frequency than F0. In the present SEEG data, we did not observe cortical FFR to first harmonic (340 Hz for *ba170*, 400 Hz for *ba200*) neither to first formant (near 800 Hz for both *stimuli*). Moreover, we observed less contacts exhibiting FFR to *ba200* than FFR to *ba170* (27 against 42). Both observations suggest a cortical phase-locking limit ranging between 200 and 340 Hz, and probably nearest to 200 Hz. Otherwise, we could simply face here a limitation of our SEEG recording system, despite the theoretical Nyquist-Shannon limit of 1024 Hz or 512 Hz for a sampling frequency of 2048 Hz.

A question remains: how this cortical FFR has remained unnoticed for all this time? Several points might allow for understanding this. Firstly, as previously said, the assumption that Speech ABR's FFR to F0 is a subcortical AEP largely relies on the FFR-to-tones literature. Yet, auditory nervous system is non-linear, so that some areas would be active for speech (seen as a weighted mixture of tone-bursts) but would not for isolated components of speech (in that view, tone-bursts), which could be linked to the behaviorally-relevant role of speech in human. Secondly, many FFR studies used animals (cat, guinea pig, marmoset monkey, amongst other) which, in addition to anatomical and functional differences of their auditory nervous system compared to humans, do not perceive speech with the same relevance as humans do. Thirdly, FFR studies initially used tones of higher frequency than 200 Hz, thus exceeding the upper frequency limit of phase-locking in A1 observed in our present study. At last, some studies showed an interpretation bias: for example, in a study where performing cryogenic cooling or removal of IC resulted in a reduced scalp-recorded FFR, authors concluded that FFR was generated by IC (Smith *et al.*, 1975); though, other generators could lie beyond IC, and require a functional IC to receive the auditory information and further exhibit FFR activity. Still, we

found one study advancing a possible contribution of A1 in generating the scalp-recorded FFR (Michelson & Vincent, 1975).

Further studies are needed to explore the non-linearity concept by focusing on the specificity of cortical FFR to speech, for example by recording SEEG activity in response to tone-bursts and speech *stimuli* of similar F0. Also, investigation of the participation of cortical FFR in generating the scalp-recorded Speech ABR could benefit from source reconstruction studies. Actually, it would be of particular interest to quantify the contribution of this cortical FFR component into the scalp-recorded response, to better understand what Speech ABR is made of, and how it can be interpreted. Finally, if participation of cortical FFR in generating scalp-recorded Speech ABR's FFR is confirmed by further studies, it would be more adapted to stop calling this response Speech ABR, but rather use a new name or an existing one, such as Speech-evoked FFR, which is neutral about the underlying generators.

Materials and Methods

First part: Stereotactic EEG.

Participants. Nine patients (ages 27-44 years, mean = 35.7, sd = 6.4; one female) with intractable epilepsy, and implanted with depth electrodes to localize their epileptic *foci* before neurosurgery, participated in the study. They were recruited when at least one of the implanted electrodes lay near the posterior Superior Temporal Gyrus on the implantation planning diagram—thus targeting A1 (Primary Auditory Cortex along Heschl's gyrus). All of them were French speakers with no self-reported hearing impairment. Each patient gave his written informed consent before recording. The protocol was approved by the regional ethic committee, and all procedures complied with the Declaration of Helsinki.

Stimuli. Two different /ba/ syllables were delivered to patients during two distinct runs. Both syllables were natural speech recorded from two native-French female speakers, and especially differed in fundamental frequency (170 or 200 Hz) and duration (180 or 200 ms, respectively). For convenience, we will refer to them as *ba170* and *ba200*. They were chosen because of their use in our previous work using topographic scalp-EEG recordings (see Bellier *et al.*, 2013 or Bellier *et al.*, 2015b for *ba170*, and Bellier *et al.*, 2015a for *ba200*).

Stimulation. Speech *stimuli* were binaurally delivered through a headset (note that it was not possible to use classical insert earphones, because of the head bandage covering ear canals), at a comfortable level of 60 dB SL (Sensation Level—above hearing threshold). The two /ba/ syllables were presented in separate runs, each consisting in 200 syllable presentations with a 930-ms SOA (Sound Onset Asynchrony—temporal distance between beginnings of two consecutive syllables) varying with a ± 20 -ms jitter. We also used alternating polarities, which consists in presenting the same amount of positive and negative syllables (here 100 of each) to minimize the potential stimulation artifact (Campbell *et al.*, 2012). Proper control of sound

level, timing and pseudo-randomization of polarities order were achieved using Presentation software (Neurobehavioral Systems).

Procedure. Patients were sitting in their hospital beds with the backrest reclined for greater comfort. The auditory threshold of the participant was first measured with an adaptive procedure using one of the *stimuli* of the main runs (the two *stimuli* were loudness-equated beforehand). Then, two consecutive runs of passive listening were proposed (*ba170*, and *ba200*), during which patients were instructed to relax and not to pay particular attention to sounds. Total duration was less than ten minutes (2' for sound level calibration task, 3' per passive listening run, plus instructions and breaks), which allowed us to adapt to both the fatigue of patients and clinical routine.

Recordings. Electrophysiological activity was recorded during the two passive-listening runs with a SystemPlus Evolution SEEG system (Micromed), at a sample rate of 2,048 Hz. Patients were implanted with ten to 16 intracranial depth electrodes (mean = 12), for a total of 92 to 181 recording sites (named contacts; mean = 127). All had at least one electrode near pSTG. We used the online reference and ground contacts as defined by clinical staff.

Data processing. SEEG data were analyzed using Elan (Aguera *et al.*, 2011) and MATLAB (The Mathworks). Raw signal was first filtered by either an 80-Hz high-pass, a 160-to-180-Hz or a 190-to-210-Hz band-pass zero-phase Butterworth filter (fifth-order). For more convenience in text and figures, resulting data are designated as HP80, BP160-180 and BP190-210. HP80 corresponds to the filter classically used in Speech ABR studies, meant to isolate sub-cortical from cortical scalp-recorded activity (Skoe & Kraus, 2010), whereas BP160-180 and BP190-210 allow to focus on fundamental frequencies of each of the /ba/ syllable (170 and 200 Hz). Then, we re-referenced the data with bipolar montages: each contact had its neighbor's signal subtracted, beginning with the deepest contact (H'1-H'2, H'2-H'3, *etc.*); this method minimizes activity from distant sources (almost identical on neighbor contacts; thus, bipolar re-referencing also minimizes a potential *stimulus* artifact), and improves spatial resolution up to 5 mm (Lachaux *et al.*, 2003; Lachaux *et al.*, 2012). A visual check allowed to ensure that no epileptic activity was present on the contacts. Filtered bipolar data were epoched in 500-ms trials (-100 to 400 ms), and all trials were averaged for both polarities, with baseline correction between -100 and 0 ms. Finally, data from opposite polarities were averaged, to obtain the intracranial Auditory-Evoked Potentials (AEP).

Statistical analysis. To identify contacts with high-frequency (superior to 80 Hz) evoked activity, we performed non-parametric Wilcoxon tests on HP80 bipolar data for each participant (to find the samples within a -100-to-400-ms epoch where the electrophysiological signal significantly differed from the -100-to-0-ms baseline; False-Rate Detection correction for multiple comparisons, with a 0.05 threshold). Then, we reduced data to a subset of contacts, by selecting those with at least two significant samples in either *ba170*- or *ba200*-evoked HP80 intracranial AEP. New Wilcoxon tests were run for these subsets of contacts in BP160-180 and BP190-210 bipolar data. Finally, for each run (*ba170* and *ba200*), we counted the number of contacts with at least two significant samples in the Wilcoxon FDR-corrected results of HP80 (all contacts), BP160-180 (contacts subset) and BP190-210 (contacts subset) bipolar data. This led us to four contact profiles, associated with a color code for Fig. 2, 3 and 4: significant activity

in HP80 only (yellow), HP80 and BP160-180 (orange), HP80 and BP190-210 (brown), and HP80, BP160-180 and BP190-210 (red).

Anatomical data. Patients underwent pre- and post-implantation anatomical MRI for clinical purpose, which we used to precisely locate electrodes and contacts in the normalized MNI space. Pre-implantation MRI was first co-registered on an anatomical T1 template in MNI, then the obtained transformation matrix was applied on post-implantation MRI using SPM8. A custom software allowed us to calculate contact coordinates from the MNI-normalized post-implantation MRI. Electrodes were represented on sagittal slice of the T1 MNI template of SPM8 in Fig. 2, and on transversal slices of (MNI-normalized) individual anatomies in Fig. 3 and 4.

Second part: Scalp EEG

Data origin. To confront intracranial AEP obtained in SEEG with classical scalp-recorded Speech ABR, we re-analyzed two datasets of our previous studies in which *Speech ABR* to *ba170* and *ba200* were acquired (see Bellier *et al.*, 2013 for *ba170*, and Bellier *et al.*, 2015a for *ba200*). In the first study, 12 normal-hearing participants (21-39 years old, mean = 27.5, sd = 6.3; 8 females) passively listened to four speech syllables (among which *ba170*) at a stimulation cadency of 3.15 Hz, while they were watching a silent movie with subtitles. Electrophysiological activity during 2,000 repetitions of each syllable in a pseudo-random order was recorded using 32 active electrodes. In the second study, 16 normal-hearing participants (14 from published study plus two additional subjects; 20-48 years old, mean = 27.9, sd = 7.6; 6 females) passively listened to *ba200* at a stimulation cadency of 1.1 Hz, while performing a visual task. Electrophysiological activity during 3,000 repetitions of *ba200* was recorded using 31 passive electrodes. In both studies, speech *stimuli* were binaurally delivered at a comfortable sound level of 50 dB SL, through E-A-RTONE 3A insert earphones (Aearo) and with alternating polarities. Electrophysiological signal was sampled at 5 kHz using a BrainAmp system (Brain Products), with reference electrode placed on the nose and impedances kept below 5 k Ω during recording.

Data processing. Data from both *ba170* and *ba200* studies underwent similar processing routine. Raw data were first cleared of the power-line radiation artifact, using notch filters on the 50 Hz and its odd harmonics. Then, resulting data went in parallel through a 3-to-20-Hz or an 80-to-1000-Hz band-pass zero-phase Butterworth filter (third- and fifth-order, respectively); this is classically done in Speech ABR studies to separate cortical “low-frequency” from subcortical “high-frequency” responses. Averaging of epoched trials (-50 to 350 ms for *ba170*, -50 to 500 ms for *ba200*) was done with rejection of those exceeding a threshold of amplitude dynamics at any electrode (80 μ V for 3-to-20-Hz data, and 100 μ V for 80-to-1000-Hz data), and with baseline correction (-50 to 0 ms for both studies). Last step to obtain Speech ABRs was an offline re-referencing of 80-to-1000-Hz data to average mastoids (known as vertical montage). In addition, Speech ABRs from both studies were filtered with a 160-to-180-Hz and a 190-to-210-Hz band-pass zero-phase Butterworth filter (fifth-order; for *ba170* and *ba200*, respectively).

Participant selection. Prior to computing grand averages, signal-to-noise ratios (SNRs) of each Speech ABR were calculated and participants with SNR below 0 dB were discarded from further analysis, resulting in the exclusion of one participant in the *ba170* study, and of two participants in the *ba200* study. Moreover, one participant in the *ba200* study shown a rare physiological artifact (Sohmer *et al.*, 1977; Kraus & Nicol, 2009), where the Post-Auricular Muscle synchronizes to the fundamental frequency of the *stimulus* or to its sub-harmonic; this was observable as huge activity on mastoid electrodes only, with SNR above 20 dB using a nose reference. This participant was also excluded, so that we kept 11 participants in the *ba170* study, and 16 participants in the *ba200* study.

Grand averaging. One major issue in averaging of Speech ABRs across participants is dephasing: since one of its components—the Frequency-Following Response (FFR), is a sustained high-frequency activity with periods of a few milliseconds (approx. 5.9 ms for *ba170*, and 5 ms for *ba200*), a phase shift as small as two milliseconds between Speech ABRs of two participants would result in a grand average with eroded FFR peaks; and such phase shifts easily happen due to inter-individual neural-timing variability or differences in subjective sound level of the *stimulus*. To overcome this issue, we re-aligned Speech ABRs of participants based on a chosen FFR peak, its latency being picked up from the narrow band-pass filtered data. Aligned Speech ABRs were finally averaged across participants (see Fig. 1, in green). Despite dephasing being far less critical regarding cortical AEPs, they were also aligned before grand averaging, based on their P50 latencies for consistency.

FFR topography. To obtain representative topographies of the FFR with minimum noise, assuming that generators of this sustained activity are constant over time, we averaged the 12 first FFR peaks based on peak latencies observed in the 160-to-180-Hz (*ba170* study) or 190-to-210-Hz (*ba200* study) band-pass-filtered grand averages (see Fig. 5). For both *ba170* and *ba200* studies, the first FFR peak has been defined as the first peak beyond 14 ms—the FFR latency (Akhoun *et al.*, 2008). Epochs of ten milliseconds were made starting from each of the 12 FFR peaks.

Acknowledgments

We thank the patients of the SEEG part of this study and the participants of our two previous scalp-EEG studies. We also thank the clinical staff from the Department of *Functional neurology and epileptology* of the Pierre Wertheimer Neurological Hospital (Lyon, France), and Philippe Albouy and Mathilde Petton for their help in SEEG data recordings. Finally, we thank Françoise Lecaigard, Romain Bouet and Jérémie Mattout for their help with data analysis. This work was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) and the LABEX CELYA (ANR-11-LABX-0060) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). LB was funded by a PhD fellowship from the Région Rhône-Alpes.

References

- Aguera P-E, Jerbi K, Caclin A, Bertrand O (2011) ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Comput Intell Neurosci* 2011:158970.
- Akhoun I, et al. (2008) The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 119(4):922–933.
- Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245(1-2):35–47.
- Anderson S, Kraus N (2010) Sensory-cognitive interaction in the neural encoding of speech in noise: a review. *J Am Acad Audiol* 21(9):575–585.
- Anderson S, Skoe E, Chandrasekaran B, Kraus N (2010) Neural timing is linked to speech perception in noise. *J Neurosci* 30(14):4922–4926.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. *J Neurosci* 32(41):14156–14164.
- Bellier L, Mazzuca M, Thai-Van H, Caclin A, Laboissière R (2013) Categorization of speech in early auditory evoked responses. *INTERSPEECH*, pp 911–915.
- Bellier L, et al. (2015a) Topographic recordings of auditory evoked potentials to speech: Subcortical and cortical responses. *Psychophysiology* 52(4):594–599.
- Bellier L, et al. (2015b) Speech Auditory Brainstem Response through hearing aid stimulation. *Hear Res* 325:49–54.
- Bidelman GM, Krishnan A (2009) Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J Neurosci* 29(42):13165–13171.
- Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.
- Bidelman GM, Krishnan A (2011) Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport* 22(5):212–216.
- Bidelman GM, Gandour JT, Krishnan A (2011a) Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn* 77(1):1–10.
- Bidelman GM, Krishnan A, Gandour JT (2011b) Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci* 33(3):530–538.
- Bidelman GM, Villafuerte JW, Moreno S, Alain C (2014) Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol Aging* 35(11):2526–2540.
- Bidelman GM (2015) Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hear Res* 323:68–80.
- Campbell T, Kerlin JR, Bishop CW, Miller LM (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear* 33(1):144–150.
- Chandrasekaran B, Kraus N (2010) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47(2):236–246.

- Clinard CG, Tremblay KL (2013) Aging degrades the neural encoding of simple and complex sounds in the human brainstem. *J Am Acad Audiol* 24(7):590–599.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 112(5):758–767.
- Gerken GM, Moushegian G, Stillman RD, Rupert AL (1975) Human frequency-following responses to monaural and binaural stimuli. *Electroencephalogr Clin Neurophysiol* 38(4):379–386.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009) Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA* 106(31):13022–13027.
- Jewett DL, Romano MN, Williston JS (1970) Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 167(3924):1517–1518.
- Jewett DL, Williston JS (1971) Auditory-evoked far fields averaged from the scalp of humans. *Brain* 94(4):681–696.
- Johnson KL, et al. (2008) Brainstem encoding of voiced consonant--vowel stop syllables. *Clin Neurophysiol* 119(11):2623–2635.
- King C, Warrier CM, Hayes E, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett* 319(2):111–115.
- Kraus N, Nicol T (2009) Auditory Evoked Potentials. *Encyclopedia of Neuroscience*, eds Binder MD, Hirokawa N, Windhorst U (Springer Berlin Heidelberg), pp 214–218.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11(8):599–605.
- Krishnan A (1999) Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiol Neurotol* 4(2):95–103.
- Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166(1-2):192–201.
- Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25(1):161–168.
- Krishnan A (2007) Frequency-Following Response. *Auditory Evoked Potentials: Basic Principles and Clinical Application*, eds Burkard RF, Eggermont JJ, Don M (Lippincott Williams & Wilkins), pp 313–333.
- Krishnan A, Gandour JT, Bidelman GM (2010) The effects of tone language experience on pitch processing in the brainstem. *J Neurolinguistics* 23(1):81–95.
- Krizman J, Marian V, Shook A, Skoe E, Kraus N (2012) Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc Natl Acad Sci USA* 109(20):7877–7881.
- Lachaux JP, Rudrauf D, Kahane P (2003) Intracranial EEG and human brain mapping. *J Physiol Paris* 97(4-6):613–628.

- Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog Neurobiol* 98(3):279–301.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. *Brain* 114(Pt 1A):139–151.
- Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P (1994) Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* 92(3):204–214.
- Lütkenhöner B, et al. (2003) Localization of primary auditory cortex in humans by magnetoencephalography. *Neuroimage* 18(1):58–66.
- Michelson RP, Vincent WR (1975) Auditory evoked frequency following responses in man. *Arch Otolaryngol* 101(1):6–10.
- Møller AR, Jannetta PJ, Jho HD (1994) Click-evoked responses from the cochlear nucleus: a study in human. *Electroencephalogr Clin Neurophysiol* 92(3):215–224.
- Møller AR (2007) Neural generators for auditory brainstem evoked potentials. *Auditory Evoked Potentials: Basic Principles and Clinical Application*, eds Burkard RF, Eggermont JJ, Don M (Lippincott Williams & Wilkins), pp 336–354.
- Moushegian G, Rupert AL, Stillman RD (1973) Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr Clin Neurophysiol* 35(6):665–667.
- Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104(40):15894–15898.
- Parkkonen L, Fujiki N, Mäkelä JP (2009) Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Hum Brain Mapp* 30(6):1772–1782.
- Perrin F, Bertrand O, Pernier J (1987) Scalp current density mapping: value and estimation from potential data. *IEEE Trans Biomed Eng* 34(4):283–288.
- Pernier J, Perrin F, Bertrand O (1988) Scalp current density fields: concept and properties. *Electroencephalogr Clin Neurophysiol* 69(4):385–389.
- Plyler PN, Ananthanarayan AK (2001) Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. *J Am Acad Audiol* 12(10):523–533.
- Russo NM, et al. (2008) Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol* 119(8):1720–1731.
- Russo N, Nicol T, Trommer B, Zecker S, Kraus N (2009) Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Dev Sci* 12(4):557–567.
- Scherg M, von Cramon D (1985) A new interpretation of the generators of BAEP waves I-V: results of a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol* 62(4):290–299.
- Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324.

- Smith JC, Marsh JT, Brown WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39(5):465–472.
- Sohmer H, Pratt H (1977) Identification and separation of acoustic frequency following responses (FFRS) in man. *Electroencephalogr Clin Neurophysiol* 42(4):493–500.
- Sohmer H, Pratt H, Kinarti R (1977) Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42(5):656–664.
- Song JH, Skoe E, Banai K, Kraus N (2011) Perception of speech in noise: neural correlates. *J Cogn Neurosci* 23(9):2268–2279.
- Steinschneider M, Nourski KV, Fishman YI (2013) Representation of speech in human auditory cortex: is it special? *Hear Res* 305:57–73.
- Stillman RD, Crow G, Moushegian G (1978) Components of the frequency-following potential in man. *Electroencephalogr Clin Neurophysiol* 44(4):438–446.
- Strait DL, Kraus N (2014) Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear Res* 308:109–121.
- Yvert B, Crouzeix A, Bertrand O, Seither-Preisler A, Pantev C (2001) Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. *Cereb Cortex* 11(5):411–423.
- Yvert B, Fischer C, Bertrand O, Pernier J (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage* 28(1):140–153.
- Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 67(3):299–317.
- Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10(4):420–422.
- Worden FG, Marsh JT (1968) Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol* 25(1):42–52.

5. Axe « Speech ABR chez le malentendant appareillé »

Les apports plébiscités de l'utilisation de la Speech ABR dans les différentes étapes de la prise en charge des patients malentendants n'ont jusque-là que fait l'objet d'études conceptuelles. Ce constat vient très probablement de la relative nouveauté de ce marqueur électrophysiologique, et surtout de la difficulté d'enregistrement de la Speech ABR chez le malentendant qui, par définition, nécessite une forte intensité de stimulation à des fréquences bien spécifiques pour percevoir et donc encoder le stimulus.

Nous avons tenté de contourner ce problème en utilisant l'appareil auditif, justement conçu pour délivrer avec grande qualité des sons de parole très fortement amplifiés sur des bandes de fréquences précises, comme stimulateur dans un protocole de Speech ABR. Cependant, l'utilisation classique d'écouteurs de type inserts dans de tels protocoles se justifie par la nécessité d'éloigner (si ce n'est d'isoler) le transducteur des électrodes d'EEG, pour éviter la transmission d'un artefact de stimulation. Aussi, les appareils auditifs disposent souvent de stratégies de sous-amplification des basses fréquences (inférieures à 500 Hz), encore naturellement perçues par les patients malentendants.

Une première étude d'ordre méthodologique a été réalisée chez le normo-entendant en contrôlant la qualité de la Speech ABR obtenue *via* appareils auditifs ainsi que l'impact potentiel d'un artefact de stimulation. Dans une seconde étape, nous avons échafaudé une étude longitudinale chez le patient malentendant, visant à mettre en évidence l'effet de l'appareillage auditif sur la plasticité neuronale du système nerveux auditif au cours des quatre premiers mois de port des appareils.

Le premier travail a donné lieu à une publication et est intimement lié à un dépôt de brevet. Le recueil de données de la seconde étude est encore en cours, et se terminera fin septembre prochain. Dans l'attente des résultats définitifs qui donneront lieu à une publication, nous avons choisi de présenter nos observations préliminaires dans le présent manuscrit, sous forme rédigée.

5.1. Appareils auditifs dans un protocole Speech ABR

L'article intitulé [Speech Auditory Brainstem Response avec stimulation par des appareils auditifs] (*Speech Auditory Brainstem Response through hearing aid stimulation*) a été publié en 2015 dans le journal à comité de lecture *Hearing Research*, numéro 325, pages 49 à 54.



Speech Auditory Brainstem Response through hearing aid stimulation



Ludovic Bellier ^{a, b, *}, Evelyne Veuillet ^{a, b, c}, Jean-François Vesson ^c, Patrick Bouchet ^{a, b}, Anne Caclin ^{a, b}, Hung Thai-Van ^{a, b, c}

^a Lyon Neuroscience Research Center, Brain Dynamics and Cognition team, INSERM U1028, CNRS UMR5292, Lyon, F-69000, France

^b University Lyon 1, Lyon, F-69000, France

^c Hospices Civils de Lyon, service d'Audiologie et Explors Orofaciales, Hôpital Edouard Herriot, Lyon, F-69000, France

ARTICLE INFO

Article history:

Received 7 November 2014

Received in revised form

26 February 2015

Accepted 5 March 2015

Available online 27 March 2015

ABSTRACT

Millions of people across the world are hearing impaired, and rely on hearing aids to improve their everyday life. Objective audiometry could optimize hearing aid fitting, and is of particular interest for non-communicative patients. Speech Auditory Brainstem Response (speech ABR), a fine electrophysiological marker of speech encoding, is presently seen as a promising candidate for implementing objective audiometry; yet, unlike lower-frequency auditory-evoked potentials (AEPs) such as cortical AEPs or auditory steady-state responses (ASSRs), aided-speech ABRs (i.e. speech ABRs through hearing aid stimulation) have almost never been recorded. This may be due to their high-frequency components requesting a high temporal precision of the stimulation. We assess here a new approach to record high-quality and artifact-free speech ABR while stimulating directly through hearing aids. In 4 normal-hearing adults, we recorded speech ABR evoked by a /ba/ syllable binaurally delivered through insert earphones for quality control or through hearing aids. To assess the presence of a potential stimulus artifact, recordings were also done in mute conditions with the exact same potential sources of stimulus artifacts as in the main runs. Hearing aid stimulation led to artifact-free speech ABR in each participant, with the same quality as when using insert earphones, as shown with signal-to-noise (SNR) measurements. Our new approach consisting in directly transmitting speech stimuli through hearing aids allowed for a perfect temporal precision mandatory in speech ABR recordings, and could thus constitute a decisive step in hearing impairment investigation and in hearing aid fitting improvement.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Hearing is a central sense in our everyday life, since human communication mainly relies on speech. Across the world, millions of hearing-impaired patients would benefit from a hearing aid, and as this population is increasing, we need new solutions to handle the situation. Objective audiometry could be one of them, if it could be implemented with an efficient and objective measure that

would give a lot of relevant information in a minimal recording time.

Speech Auditory Brainstem Response (speech ABR) is an electrophysiological marker of speech encoding, which provides precise spectro-temporal information on auditory processing at the brainstem level (Skoe and Kraus, 2010). Being seen as a direct window on what information enters the auditory central nervous system, speech ABR mimics so well its stimulus that when playing it back as a sound, the stimulus can be understood (Galbraith et al., 1995). As two recent articles suggested (Anderson and Kraus, 2013; Dajani et al., 2013), its intrinsic properties make it a good candidate to facilitate hearing aid selection, fitting, and evaluation of performance. Furthermore, using speech ABR as an objective audiometric tool would be of particular interest in non- or poorly-communicative hearing-impaired patients (e.g., young children, Down syndrome, dementia).

To record speech ABR, speech syllables are usually delivered by insert earphones whereby the sound is conveyed to the

Abbreviations: Speech ABR, speech auditory brainstem response; AEP, auditory-evoked potential; ASSR, auditory steady-state response; SNR, signal-to-noise; F0, fundamental frequency; dB SPL, decibel sound pressure level; dB HL, decibel hearing level; EEG, electroencephalography; RMS, root mean square; DSS, denoising source separation; ERP, event-related potential; FFR, frequency-following response

* Corresponding author. Lyon Neuroscience Research Center, Inserm U1028/CNRS UMR 5292/Université Claude Bernard Lyon 1, Brain Dynamics and Cognition team, Centre Hospitalier le Vinatier, 95 boulevard Pinel, F-69500, Bron, France. Tel.: +33 (0)4 72 13 89 05.

E-mail address: ludovic.bellier@inserm.fr (L. Bellier).

<http://dx.doi.org/10.1016/j.heares.2015.03.004>

0378-5955/© 2015 Elsevier B.V. All rights reserved.

participant's ears via plastic tubes, thus minimizing risks of stimulus artifact (Akhoun et al., 2008). However, such systems are somewhat limited in their intensity range. Stimulation through hearing aids would be far more ecological and powerful in the hearing-impaired, allowing the delivery of specific frequencies at very-high intensity (up to 130 dB SPL) and the assessment of particular settings for each hearing loss. However, it could fail to give proper speech ABR, because of stimulus artifact or inadequate precision of timing.

Previous attempts of using electrophysiological markers in hearing aid healthcare have been made, with Click ABR (Kießling, 1983), Auditory Steady-State Response (ASSR; Picton et al., 1998) or cortical Auditory-Evoked Potentials (cAEPs; Billings et al., 2007), although none has led to a wide use in hearing aid healthcare. This might be due to their relatively low balance between relevance of information provided and recording time. In any case, what interests us in the present methodological paper is not relevance but feasibility: ASSR and cAEPs have already been recorded through hearing aid stimulation, but these activities are from lower frequencies than speech ABR, thus less demanding for a perfect hearing aid timing. Regarding ASSR, carrier frequencies can be as high as the higher frequencies present in speech stimuli, but the electrophysiological signal pertains to the low-frequency amplitude modulation and is thus less likely to be altered by a few-millisecond temporal jitter.

To the best of our knowledge, up to now, only one paper (Anderson and Kraus, 2013) reported preliminary speech ABR data in a single participant obtained with open-field stimulation amplified by hearing aids. Here, we assess more comprehensively the feasibility of using hearing aid stimulation to record artifact-free speech ABR. One major novelty here is that we chose to directly transmit the stimulus to hearing aids using a wireless connection, which allows (1) to avoid latency jitters caused by head movements and (2) to test mute conditions to properly control for the absence of stimulus artifact (see Methods). As obtaining speech ABR for individual participants is of particular interest for hearing aid benefit assessment, we also assessed the use of an advanced multi-channel-data processing method to improve signal-to-noise (SNR).

2. Material and methods

2.1. Participants

Four healthy adults (aged 22–25 years; 3 women) participated in the study. All were normal-hearing (with air conduction hearing thresholds measured below 20 dB HL between 250 and 8000 Hz with pure tone audiometry), native French speakers, and none had known language, psychiatric, or neurologic impairment. Each participant gave his/her written informed consent. All experimental procedures were approved by the “Comité de Protection des Personnes Lyon SUD-EST II” (#2008-017-2).

2.2. Stimulus

Speech stimulus was a 180-ms natural /ba/ syllable (Bellier et al., 2013), recorded from a French female voice. It was composed of a 45-ms voicing followed by a 135-ms vowel (Fig. 1). Fundamental frequency (F0) of the vowel/a/was slightly varying around 170 Hz, and the four first formants were respectively centered around 740, 1360, 2785 and 4310 Hz. The most energetic frequency during the vowel was F0.

2.3. Transducers

We recorded electrophysiological activity in response to the /ba/ stimulus in three runs using different transducers: insert earphones or hearing aids with two different settings (see below). Two additional runs were recorded with insert earphones and hearing aids but muted: transducers were in the same position as in normal runs and also played the /ba/ stimulus, but foam ear tips were filled with silicone and removed from ear canals, making the stimulus inaudible. The order of these five runs was randomized between participants.

Hearing aids were a pair of Siemens Motion 501 SX XCEL (Munich, Germany), and insert earphones were the Aearo E-ARTONE 3A (Indianapolis, USA) – the most frequently used transducer in speech ABR studies. Both transducers were coupled with foam ear tips, to avoid leakage of low frequencies from ear canals. Two different settings were used for hearing aids, sharing all parameters but gains to obtain two different waveforms. Recordings of each transducer output were made using a G.R.A.S. 43AG ear simulator (Holte, Denmark), and are shown in Fig. 1 (average of 80 recorded stimuli). Stimuli were played directly through the hearing aids using the wireless transmission program, or through the insert earphones. Timing of stimulation through hearing aids and the wireless connection was highly reproducible (no jitter), but with a fixed delay of 62.2 ms (corrected in Fig. 1). For each transducer, sound level was measured with a 2cc coupler linked to the 43AG ear simulator and mounted on a Brüel & Kjær 2239A sound level meter (Nærum, Denmark), and sound card output level was adjusted to reach 80 dB SPL A-weighted.

2.4. Procedure

Participants sat in an armchair located in an electromagnetically-shielded soundproofed room, and watched a subtitled movie without soundtrack during the five runs. In each run, 3000 /ba/ syllables were binaurally delivered at a sound level of 80 dB SPL, with a rate of 2.78 per second, in alternating polarities (1500 each), using Presentation software (Neurobehavioral Systems; Berkeley, USA). Each run lasted 18 min, and was followed by a 5-minute break during which the transducers were changed.

2.5. Electrophysiological recordings and data processing

Speech ABRs were recorded using a 32-channel BrainAmp EEG system with ActiCAP active electrodes (Brain Products; Munich, Germany), at a sampling rate of 5000 Hz (Bellier et al., 2015). Reference was placed on the nose, and ground at AFz. Impedances were kept below 5 k Ω during all the procedure. Data were analyzed using Elan (Aguera et al., 2011) and MATLAB (The MathWorks; Natick, USA). For each run, raw data were first filtered with notch filters to remove the 50-Hz power line radiation artifact and its odd harmonics, then with an 80–1000-Hz Butterworth bandpass filter (fifth order) to isolate sub-cortical activity. Filtered data were epoched in 400-ms trials between –50 and 350 ms, and trials exceeding 100 μ V of amplitude dynamics at any electrode were rejected. Accepted trials were averaged (1403 ± 50 out of 1500 trials per polarity; mean \pm SD), and baseline correction was applied (–50 to 0 ms). Speech ABRs were obtained using a vertical montage by offline re-referencing the vertex to the average of mastoids (Cz–(M1 + M2)/2). We computed both envelope and spectral FFRs, respectively by averaging responses from opposite polarities or by averaging the response from one polarity and the negative of the response from the opposite polarity. Envelope FFR typically follows

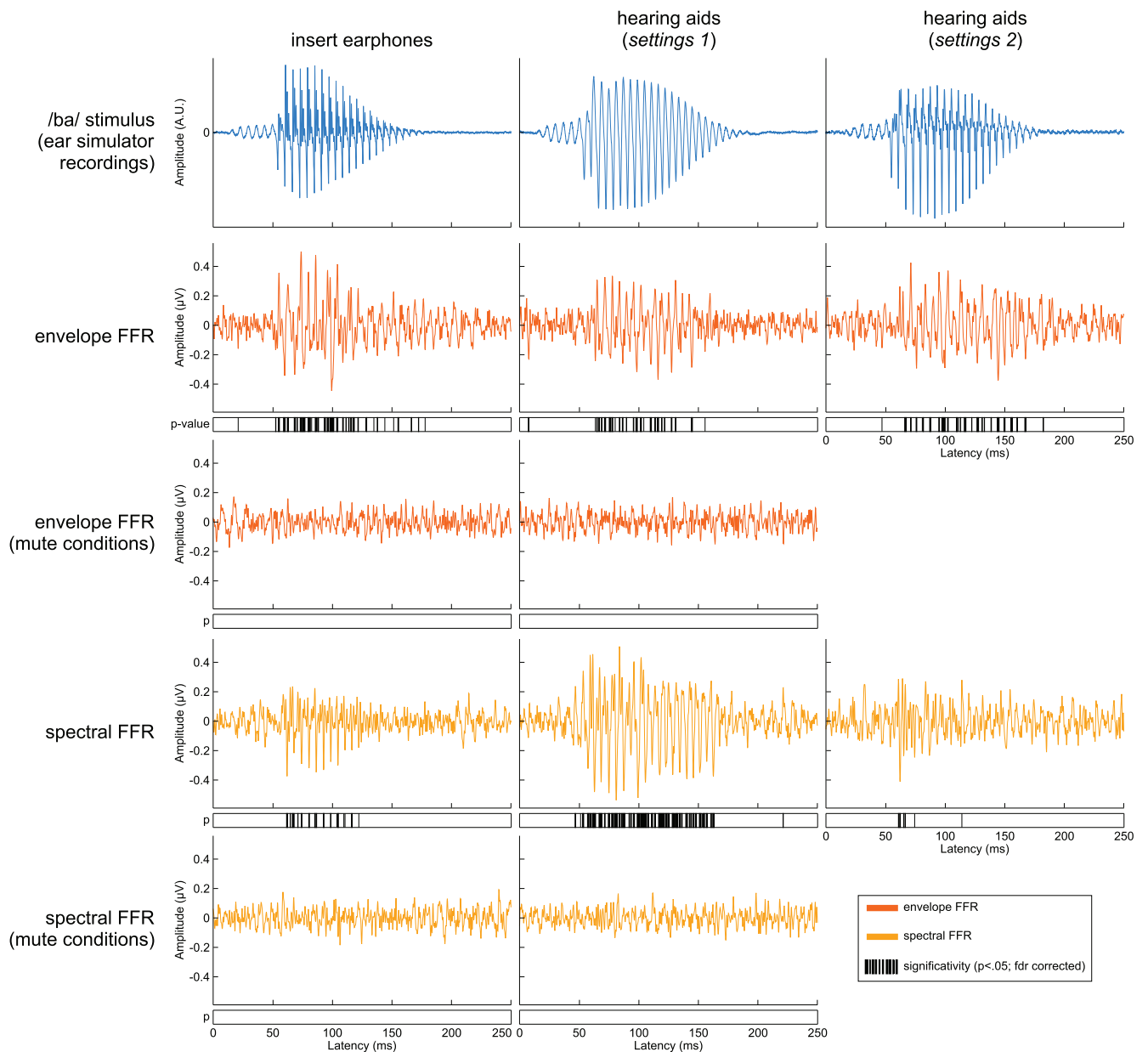


Fig. 1. Speech ABRs for each transducer in a single participant, with a vertical montage. Acoustical restitution of the /ba/ syllable is shown line 1 (blue), for each transducer, as recorded by an ear simulator. Speech ABRs elicited by these stimuli are presented for envelope FFR and spectral FFR on lines 2 (orange) and 4 (yellow), respectively; significant samples (Wilcoxon tests; FDR-corrected for multiple tests with a threshold of 0.05) are represented below each plot. In mute conditions, speech ABRs are shown for envelope FFR (line 3, orange) and spectral FFR (line 5, yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the stimulus fundamental frequency and minimizes the potential stimulus artifact (Campbell et al., 2012), while spectral FFR mimics the spectral content of the stimulus with a maximum of energy corresponding to the stimulus spectral maximum but also reveals the potential stimulus artifact (Aiken and Picton, 2008). Finally, Fast-Fourier Transforms were computed for vowels of each transducer's stimulus (50–170 ms), and for FFR of each DSS-processed response for each participant (60–180 ms, to accommodate for the ~10 ms delay of the response with respect to the stimulus), before averaging across the four participants.

2.6. Signal-to-Noise (SNR) calculation

Speech ABR quality was estimated with SNR, using the formula: $20 \cdot \log_{10}((E_{XY} - E_Y)/E_Y)$; with E_Y being the energy in the noise

window (Root Mean Square (RMS) of the –50 to 0-ms speech ABR portion) and E_{XY} being the energy of the signal + noise window (RMS of the 60–110-ms speech ABR portion). SNR is expressed in decibels (dB), and an SNR equal to zero means that E_{XY} is worth the double of E_Y .

2.7. Denoising Source Sepan (DSS)

We used DSS to improve SNR by taking advantage from the multichannel-recording setup. DSS consists in designing spatial filters to decompose signal into as many components as electrodes (here 32) using a criterion of stimulus-evoked reproducibility (de Cheveigné and Simon, 2008; Särelä and Valpola, 2005). For each run and FFR type, we calculated the SNR for the best component (representing the most reproducible part of the ERP) and used its

weights array (corresponding to the topography of the component) to re-reference EEG data prior to statistical analysis.

2.8. Statistical analysis of signal emergence

To assess whether a speech ABR signal emerged with respect to a baseline period, non-parametric paired-comparison Wilcoxon tests were conducted on filtered and re-referenced EEG data (vertical montage or virtual channel corresponding to the best-DSS-component weights array). As Wilcoxon tests were conducted for each sample, we computed the False Detection Rate (fdr) statistics with a threshold of 0.05 to correct for multiple comparisons.

3. Results

Speech ABRs of a single participant are presented in Fig. 1. When participant heard the /ba/ stimulus (Fig. 1 lines 2 and 4), we obtained significant envelope and spectral FFRs regardless of the transducer. Moreover, in mute conditions (Fig. 1 lines 3 and 5), no significant activity was found. If a stimulus artifact was present, it should be revealed by spectral-FFR-like processing (see Methods), which was not the case here. The same pattern of results was indeed observed for all four participants, as shown Fig. 2 through the quantification of speech ABRs emergence relative to baseline, in terms of the number of significant samples in each participant. In all audible conditions (Fig. 2 line 1), significant speech ABRs were detected, while no significant sample was found in mute conditions (Fig. 2 line 2), excepted for DSS processing of spectral FFR with two significant samples found in two participants. These results show that it is possible to record genuine electrophysiological activity (speech ABRs) with hearing aid stimulation.

Signal-to-noise ratios (SNRs) for each participant are shown in Fig. 3. In audible conditions (Fig. 3 line 1), mean SNRs are all above 3 dB, whereas in mute conditions (Fig. 3 line 2), they are all below -4.4 dB; this means high-quality speech ABRs were obtained whatever the tested transducer, while no stimulus artifact did emerge from noise in mute conditions, even for spectral FFR. SNRs between the three transducers in the audible conditions are comparable: hearing aid stimulation leads to high-quality speech ABRs, and does as good as insert earphones for the same stimulation level

(here around 80 dB SPL A). For all transducers and FFR types, in audible conditions (Fig. 3 line 1), DSS improved SNR from about 3.3 dB; however, in mute conditions (Fig. 3 line 2), it did not consistently improve SNR, revealing that benefits of DSS are only effective when genuine signal is present in the data.

Spectra of stimuli vowels, along with grand-average spectra of the corresponding DSS-processed envelope and spectral FFRs are shown in Fig. 4. In the stimuli spectra, maxima of energy were around the 170 Hz fundamental frequency (F0), which was especially observable for hearing aids with Settings 1 (see also Fig. 1 for its corollary in time domain). These stimuli spectral maxima were faithfully mirrored in the corresponding spectral FFRs, as described in the literature (see Figures 4 and 6 in Aiken and Picton, 2008), with the highest energy at F0 especially for hearing aids with Settings 1. As expected, envelope FFRs also had a maximum of energy at F0, but its amplitude was not determined by the energy at F0 in the stimulus spectrum.

In terms of spectral content, the two closest stimuli were hearing aids with Settings 2 and insert earphones (Fig. 4, left panel). Interestingly, these two stimuli led to a similar quality of envelope and spectral FFR as assessed with number of significant samples (Fig. 2, left and right panels) and SNR (Fig. 3, left and right panels), whereas using hearing aids with Settings 1 the spectral FFR was strengthened because of the increased energy at F0. These observations corroborated a strong dependence of speech ABR to its eliciting stimulus spectral content and temporal fine structure.

4. Discussion

Auditory stimulation through hearing aids led to significant high-quality speech ABRs, exempt of any detectable stimulus artifact. Collecting this electrophysiological marker in a single participant is thus fully viable, and can be further improved with our present methodology using advanced multichannel data processing methods like DSS (see also Bellier et al., 2015).

Results obtained here with a pair of Siemens Motion 501 SX XCEL using two different settings could be generalized to every hearing aid, taking some simple precautions: checking (1) for a reproducible timing (no jitter), (2) for a faithful “in-ear” acoustic signal, and (3) for the absence of stimulus artifact as done here

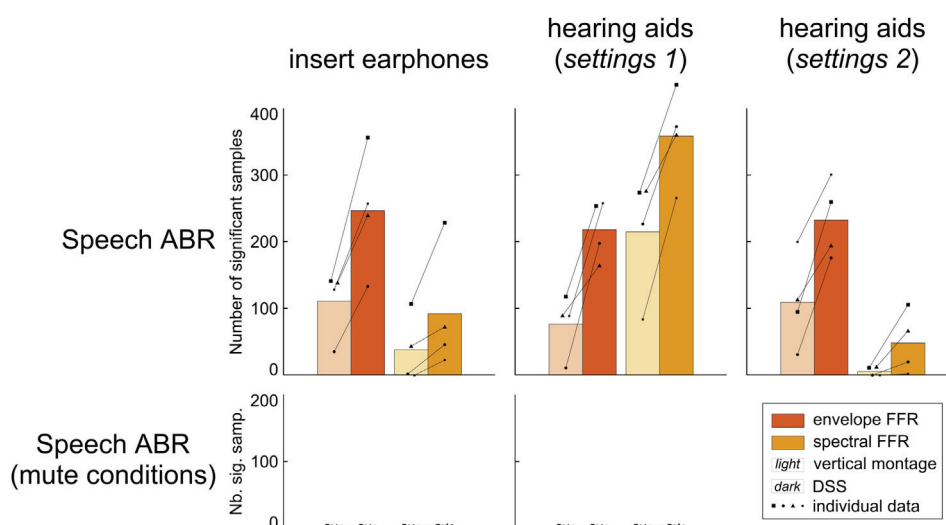


Fig. 2. Number of significant samples (over the entire analysis time window: -50 to 350 ms) in speech ABRs for each transducer, FFR type (envelope or spectral) and processing (vertical montage or DSS). Bars represent mean values of the four participants, and dots of different shapes represent individual data. Colors differentiate FFR types (as in Fig. 1). Significant samples in speech ABRs obtained with a vertical montage are shown in light colors, while dark colors are used for speech ABRs obtained using DSS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

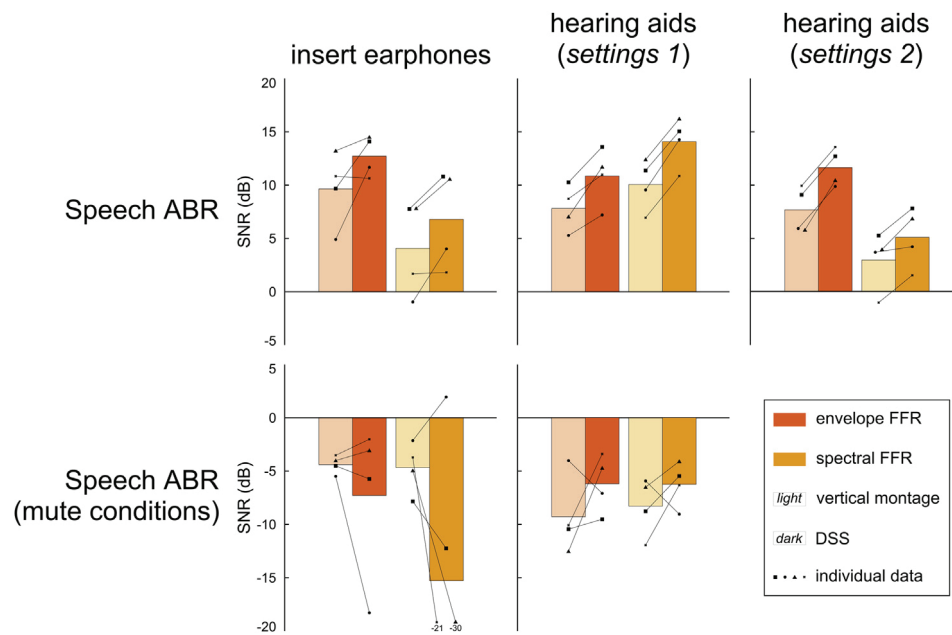


Fig. 3. Signal-to-noise ratios (SNRs, in decibels, dB) of speech ABRs for each transducer, FFR type and processing. Color codes match those of Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using mute conditions. Reproducible timing is crucial, in that a jitter of a few milliseconds would completely erase FFR by counter-phasing peaks between single trials. In that view, aided-speech ABR is more demanding than aided-cAEPs or ASSR. Acoustic signal quality is of importance, because a distorted signal could lead to unrecognizable speech ABRs; specifically, presence of energy at the fundamental frequency is critical to get a proper FFR. Consequently, we did not use “open ear tips” to avoid low-frequency leakage from ear canals. Presence of energy at the fundamental frequency should be assessed after application of hearing aid gains for hearing impaired participants. The absence of stimulus artifact, finally, is highly desirable, because it makes the spectral FFR exploitable, which is of particular interest since a recent paper pointed out that its balance with envelope FFR was changed in hearing impaired participants (Anderson et al., 2013). If a stimulus artifact exists, it can however be minimized using envelope FFR (summing opposite polarities; see

Campbell et al. (2012)). The proper control for a stimulus artifact using mute conditions provides a novel insight to the literature of aided-AEPs. To our knowledge, such a control has never been reported for aided-ASSR, which could be done using mute conditions as used here. In the present study, using a direct wireless transmission of signal to hearing aids limited the risk of recording any stimulation artifact at the stimulus frequency, as it was not conveyed in an analog fashion anywhere close to the recording electrodes.

As just said, the stimulation strategy chosen here was to transmit the stimulus directly to hearing aids, rather than playing it in open field with speakers then amplifying it with hearing aids. Indeed, with this last strategy, head movements of the participant could induce temporal jitter and reduce speech ABR amplitude; moreover, it would have been complicated to assess the presence of a stimulus artifact using mute conditions, as an 80 dB SPL stimulus in open field is hard to mute.

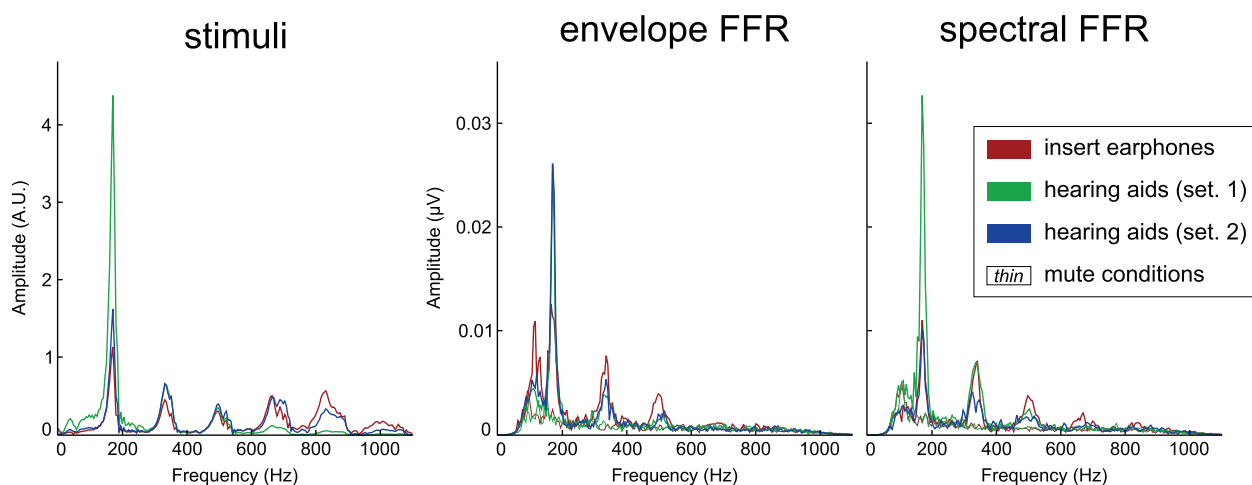


Fig. 4. Spectra of/a/vowels as delivered by transducers, and grand-average spectra of corresponding DSS-processed FFRs. Colors differentiate transducers, and thin curves represent mute conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The present study thus showed that it is possible to record genuine speech ABRs, and assess properly both envelope and spectral FFRs, by delivering speech stimulus directly through hearing aids. The results open new perspectives for improved hearing aid fitting and benefit assessment, using the extensive information provided by this electrophysiological marker.

Contributors

LB, EV, JFV, AC and HTV designed the study. LB, PB and JFV set the apparatus up. LB collected the data. LB, AC and HTV analyzed the data and wrote the paper. All authors revised the paper and approved the final version.

Conflict of interest

None of the authors had any personal or financial interest in Siemens company.

Acknowledgments

The authors thank Mikael Ménard and Pascal Boulud (Siemens Audiology France) for providing hearing aids and related equipment, and for their technical help. They also thank Alain de Cheveigné for his help on how to use DSS.

This work was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) and the LABEX CELYA (ANR-11-LABX-0060) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). LB was funded by a PhD fellowship from the Région Rhône-Alpes.

References

- Aguera, P.-E., Jerbi, K., Caclin, A., Bertrand, O., 2011. ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Comput. Intell. Neurosci.* 158970.
- Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.
- Akhoun, I., Moulin, A., Jeanvoine, A., Ménard, M., Buret, F., Vollaire, C., Scorretti, R., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. *J. Neurosci. Methods* 175, 196–205.
- Anderson, S., Kraus, N., 2013. The potential role of the cABR in assessment and management of hearing impairment. *Int. J. Otolaryngol.* 604729.
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., Kraus, N., 2013. Effects of hearing loss on the subcortical representation of speech cues. *J. Acoust. Soc. Am.* 133, 3030–3038.
- Bellier, L., Mazzuca, M., Thai-Van, H., Caclin, A., Laboissière, R., 2013. Categorization of speech in early auditory evoked responses. *Conf. InterSpeech Lyon* 837–841.
- Bellier, L., Bouchet, P., Jeanvoine, A., Valentin, O., Thai-Van, H., Caclin, A., 2015. Topographic recordings of auditory evoked potentials to speech: sub-cortical and cortical responses. *Psychophysiology* 52, 594–599.
- Billings, C.J., Tremblay, K.L., Souza, P.E., Binns, M.A., 2007. Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials. *Audiol. Neurotol.* 12, 234–246.
- Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear. Hear.* 33, 144–150.
- Dajani, H.R., Heffernan, B.P., Giguere, C., 2013. Improving hearing aid fitting using the speech-evoked auditory brainstem response. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2812–2815.
- De Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. *J. Neurosci. Methods* 171, 331–339.
- Galbraith, G.C., Arbagey, P.W., Branski, R., Comerchi, N., Rector, P.M., 1995. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6, 2363–2367.
- Kiessling, J., 1983. Clinical experience in hearing-aid adjustment by means of BER amplitudes. *Arch. Oto-Rhino-Laryngol.* 238, 233–240.
- Picton, T.W., Durieux-Smith, A., Champagne, S.C., Whittingham, J., Moran, L.M., Giguère, C., Beauregard, Y., 1998. Objective evaluation of aided thresholds using auditory steady-state responses. *J. Am. Acad. Audiol.* 9, 315–331.
- Särelä, J., Valpola, H., 2005. Denoising source separation. *J. Mach. Learn. Res.* 6, 233–272.
- Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. *Ear. Hear.* 31, 302–324.

5.2. Plasticité neuronale chez le malentendant appareillé

Les résultats préliminaires de cette étude longitudinale sont ici présentés sous forme rédigée.

Plasticité neuronale chez le malentendant appareillé

Introduction

Les troubles de l'audition affectent des millions d'individus à travers le monde, et sont bien souvent la cause d'une dégradation de la qualité de vie sociale associée à un mal-être (Carlsson *et al.*, 2011 ; Ciorba *et al.*, 2012 ; Warner-Czyz *et al.*, 2015). La presbyacousie, ou vieillissement physiologique de l'oreille, représente une grande proportion de ces troubles (Gates & Mills, 2005). Aussi, lorsqu'il n'est pas pris en charge, le patient presbyacousique se renferme peu à peu sur lui-même, d'abord gêné puis lassé et frustré d'avoir à faire répéter ses interlocuteurs. Et les effets négatifs d'une surdité non traitée ne s'arrêtent pas là : un nombre croissant d'études met en lumière les liens entre presbyacousie, isolation sociale et maladie d'Alzheimer (Wilson *et al.*, 2007 ; Lin *et al.*, 2011 ; Albers, 2012 ; Bakhos *et al.*, 2015 ; Panza *et al.*, 2015).

Classée dans les surdités neurosensorielles, la presbyacousie se manifeste par une atteinte de la cochlée (au niveau des cellules ciliées et de la strie vasculaire) et des neurones afférents du nerf auditif (Sakaguchi *et al.*, 2009 ; Bao & Ohlemiller, 2010 ; Yamasoba *et al.*, 2013 ; Roth, 2015 ; Ouda *et al.*, 2015 ; Wong & Ryan, 2015). En résultent une baisse de l'audibilité dans les hautes fréquences acoustiques, qui jouent un rôle primordial dans l'intelligibilité de la parole, et une résolution fréquentielle dégradée. Dans bien des cas, les aides auditives (aussi appelées appareils ou prothèses) permettent de compenser ces déficits, en amplifiant sélectivement les fréquences perdues.

Au niveau comportemental, de nombreuses études ont attesté et parfois contesté les bénéfices de l'utilisation d'aides auditives sur la discrimination fréquentielle (Gabriel *et al.*, 2006 ; Kluk & Moore, 2006 ; Thai-Van *et al.*, 2010), la discrimination en intensité et la perception de la sonie (Robinson & Gatehouse, 1995 ; Philibert *et al.*, 2002, 2005 ; Munro & Trotter, 2006 ; Munro & Merrett, 2013), et la perception de parole (Gatehouse, 1992 ; Köbler & Rosenhall, 2002 ; Munro & Lutman, 2003 ; Lavie *et al.*, 2015). Notamment, une étude a montré un effet de l'appareillage auditif sur l'identification de l'indice phonétique de voisement, avec une amélioration de la pente de la sigmoïde de perception catégorielle au bout de 4 mois d'utilisation, et ce seulement pour l'oreille droite (Philibert *et al.*, 2003). Ces résultats positifs, provenant d'un changement prolongé dans la qualité de l'information spectrale apportée au système nerveux auditif, sont la

manifestation de l'effet d'acclimatation auditive, parfois observable après seulement quelques semaines d'appareillage (Arlinger *et al.*, 1996 ; Munro, 2008).

En comparaison, très peu d'études se sont intéressées aux changements électrophysiologiques conséquents à un appareillage auditif. Au niveau sous-cortical, l'une d'elles a mis en évidence un raccourcissement de la latence de l'onde V des Potentiels Evoqués Auditifs Précoces (PEAP, ou Click ABR) seulement pour l'oreille droite, au cours des 6 premiers mois d'appareillage (Philibert *et al.*, 2005) ; ce résultat était interprété comme la preuve d'une resynchronisation neuronale de l'encodage des hautes fréquences nouvellement apportées par les aides auditives, les clics étant liés à la sensibilité aux fréquences de 2 à 4 kHz (Stapells & Oates, 1997). Une autre étude a pointé une augmentation de l'amplitude de l'onde V des PEAP (plus exactement de la différence d'amplitude entre l'onde V et SN₁₀, son rebond négatif) chez des patients malentendants utilisant une aide auditive monaurale depuis au moins 2 ans (Munro *et al.*, 2007). Cependant, une troisième étude n'a pas retrouvé ces résultats après 3 mois d'appareillage (Dawes *et al.*, 2013). Enfin, au niveau cortical, une étude a montré une augmentation de l'amplitude de l'onde P2 des PEA corticaux seulement pour un son à 2 kHz délivré à l'oreille appareillée, observée chez des patients utilisant une aide auditive depuis au moins 5 ans (Bertoli *et al.*, 2011) ; tandis qu'une seconde étude n'a pas retrouvé d'effet sur ces réponses corticales tardives après 3 mois d'appareillage (Dawes *et al.*, 2014).

L'ensemble de ces résultats suggère très clairement l'existence d'améliorations comportementales induites par l'utilisation d'aides auditives, en parallèle d'une réorganisation fonctionnelle des voies nerveuses auditives, bien que la durée d'utilisation requise et la latéralité des effets soient sujettes à discussion. Cependant, la non-linéarité du système auditif humain fait qu'un effet d'acclimatation auditive peut ne se manifester pleinement qu'avec une stimulation par des sons de parole, et non avec des sons plus simples (tons purs, clics). Or, l'effet de l'utilisation d'aides auditives sur l'encodage neuronal des sons de parole n'a jamais été testé.

La Speech Auditory Brainstem Response (Speech ABR) constitue justement un marqueur électrophysiologique de l'encodage des sons de parole dans le système nerveux auditif humain (Skoe & Kraus, 2010). Si elle n'a jamais été enregistrée chez le patient malentendant appareillé, en partie à cause de limitations d'ordre technique, une récente étude a révélé une surreprésentation des basses fréquences dans les Speech ABR de patients malentendants atteints de surdité neurosensorielle (Anderson *et al.*, 2013), suggérant un effet d'acclimatation à la privation auditive : recevant moins d'afférences depuis la région basale de la cochlée codant pour les hautes fréquences, les voies auditives se réorganiseraient alors en faveur de l'encodage des basses fréquences mieux conservées. En ce qui concerne les hautes fréquences, une étude antérieure combinant Speech ABR et tâche d'identification de syllabes a montré que l'encodage neuronal du deuxième

formant, situé dans le spectre entre 800 et 2500 Hz et associé à la perception du lieu d'articulation des syllabes, était dégradé chez le malentendant ; ce résultat était lié chez certains patients à une diminution des capacités d'identification phonémique du lieu d'articulation, établissant de ce fait un lien entre perception des consonnes et encodage neuronal des hautes fréquences (Plyler & Ananthanaryan, 2001).

Dans la présente étude, notre hypothèse principale est que l'utilisation d'aides auditives permet de contrebalancer les dégradations observées dans l'encodage neuronal et la perception des sons de parole, et ce au cours des premiers mois d'appareillage. En d'autres termes, nous questionnons l'existence d'un effet d'acclimatation sur l'encodage neuronal et la perception de parole du patient malentendant appareillé. Notre hypothèse secondaire est que cet effet s'observera d'autant plus pour les sons de parole dont les indices phonétiques étaient d'autant moins perçus avant appareillage.

Pour tester ces hypothèses, nous avons conçu une étude longitudinale avec 5 sessions s'étalant sur 4 mois, chez un groupe de patients malentendants appareillés pour la première fois, et chez un groupe contrôle d'individus normo-entendants permettant de nous assurer de l'absence d'un éventuel effet d'entraînement. A chaque session, les capacités d'identification phonémique de syllabes s'opposant en lieu d'articulation (/ba/-/da/) et en nasalité (/ba/-/ma/) étaient d'abord évaluées, puis un enregistrement électrophysiologique des Speech ABR et des PEA corticaux évoqués par les mêmes syllabes était effectué. Aussi, et de façon inédite, les stimuli ont été délivrés aux oreilles des participants *via* des appareils auditifs réglés selon les besoins de chacun (réglages de l'audioprothésiste pour les malentendants, et calculés pour une audiométrie normale pour les normo-entendants), et recevant le signal acoustique par alimentation électrique directe afin de garantir une synchronisation parfaite avec le système d'électroencéphalographie (EEG). Notons que nos précédents travaux ont servi de base méthodologique pour la tâche comportementale d'identification phonémique (Bellier *et al.*, 2013), l'enregistrement optimal de la Speech ABR et des PEA corticaux (Bellier *et al.*, 2015a), et la mise au point de la stratégie de stimulation *via* les aides auditives (Bellier *et al.*, 2015b).

Résultats

Nous présentons dans cette partie les résultats comportementaux et électrophysiologiques issus d'une première série d'analyses de ces données multifactorielles. Les différents facteurs susceptibles d'influencer les variables dépendantes sont le groupe (10 patients et 4 contrôles), la session (t_0 , t_1 , t_2 , t_3 ou t_4), et le *continuum* (/ba/-/da/ ou /ba/-/ma/) pour les données comportementales ou le stimulus (*ba*, *da30*, *da70*, *ma30* ou *ma70*) pour les données électrophysiologiques. Nous avons d'abord analysé l'effet des facteurs inter-sujets (groupe) et intra-sujet (*continuum* et stimulus) précités avec une ANOVA à mesures répétées, pour les différentes mesures comportementales et électrophysiologiques d'intérêt.

Ensuite, pour potentialiser la mise en évidence d'éventuels effets longitudinaux (évolution des mesures au cours des sessions), nous avons comparé chaque variable dépendante entre t_0 et t_4 pour le groupe de 10 patients malentendants, en utilisant un test non paramétrique de Wilcoxon. Puis, pour les tests significatifs, nous avons comparé les groupes *malentendants* et *normo-entendants* sur la proportion de participants manifestant l'effet, en utilisant un test de Khi-deux : une différence significative entre les groupes suggère alors que l'effet longitudinal observé est bien lié à l'utilisation des aides auditives réglées de façon appropriées, et non à un quelconque effet non-spécifique d'entraînement à la tâche ou au matériel expérimental. Notons que cette stratégie d'analyse a été appliquée à l'intégralité des variables dépendantes comportementales et électrophysiologiques, et que sont décrits ici l'ensemble des résultats significatifs.

Un premier regard sur les données et une comparaison des différentes variables entre les groupes permet de repérer les déficits comportementaux ou électrophysiologiques des patients malentendants, pour ensuite pouvoir cibler leur évolution dans l'hypothèse d'une compensation de ces déficits induite par l'appareillage auditif. Précisons que cette stratégie est une première approche qui n'exclut pas l'étude ultérieure de variables ne montrant pas de différence entre patients et contrôles, celles-ci pouvant aussi évoluer au cours de l'utilisation des aides auditives.

Identification phonémique (données comportementales).

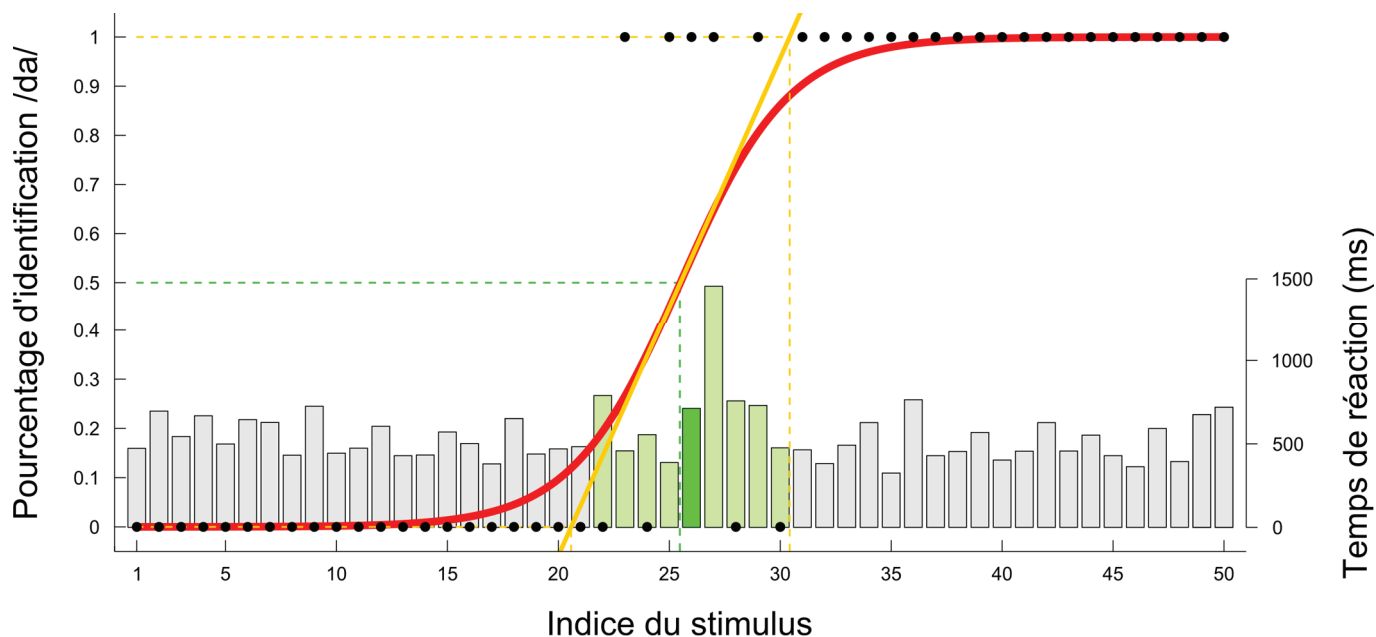


Figure 1. Courbe de perception catégorielle du patient P10 pour le *continuum* /ba-/da/ (50 stimuli en abscisse) à *t2*. Voir détails dans le texte ci-dessous.

Avant de partir dans l’exploration des données d’identification phonémique, une explication des différentes mesures est souhaitable (Figure 1). Ainsi, les points noirs représentent les réponses du participant (0 correspond à une réponse /ba/, 1 à une réponse /da/). La courbe sigmoïde (en rouge) est obtenue par régression logistique à partir de ces réponses. La projection du point de l’axe des ordonnées valant 50% d’identification /da/ sur cette sigmoïde (ligne verte pointillée horizontale), puis du point de la sigmoïde ainsi obtenu sur l’axe des abscisses (verticale), donne l’indice du stimulus à la **frontière** catégorielle (ici, 26). Concernant la pente de la sigmoïde, qui informe de la précision catégorielle, nous avons calculé une valeur associée et bornée entre 1 et 50, correspondant respectivement à des pentes valant l’infini et 0. Pour cela, nous avons d’abord tracé la tangente de la sigmoïde à la frontière (droite orangée en trait plein), puis obtenu les indices aux intersections de cette tangente avec les droites $y = 0$ et $y = 1$ (traits orangés pointillés), et enfin calculé la différence absolue de ces indices (ici, 10) ; comme cette valeur représente le nombre de pas dans le *continuum* séparant les deux catégories phonémiques en quelque sorte non-ambigües, nous l’appellerons « **zone d’ambiguïté** ». Deux dernières mesures étaient relevées sur les temps de réaction (barres grises et vertes avec ordonnée à droite) : le **temps de réponse** moyen pour les 9 syllabes centrées sur la frontière (barres vertes), et le temps de réponse moyen pour les 41 syllabes restantes (barres grises).

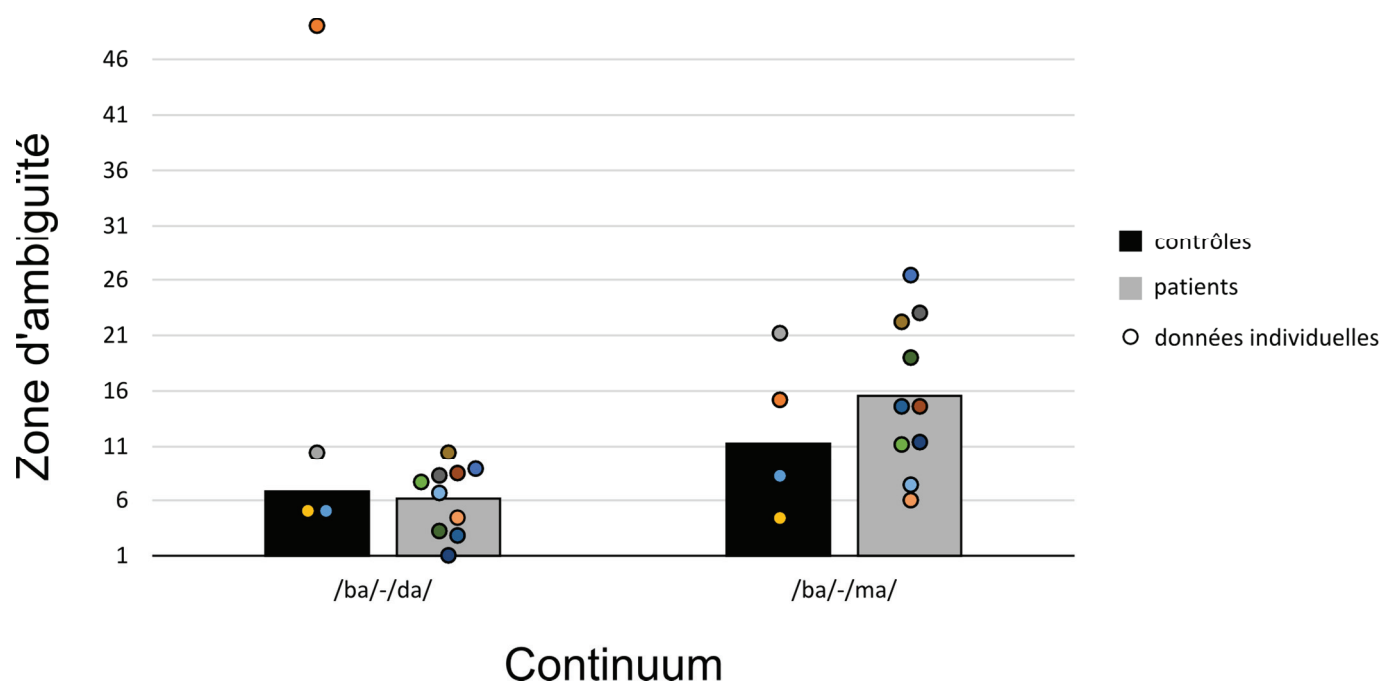


Figure 2. Taille de la zone d'ambiguïté (nombre de pas sur le *continuum* de 50 stimuli) en fonction du *continuum* et du groupe, quelle que soit la session. Les points colorés représentent les valeurs individuelles moyennes toutes sessions confondues. Notons que le point orange en haut à gauche de la figure correspond aux performances moyennes de C02 sur l'identification du *continuum* /ba-/da/, anormalement mauvaises par rapport à l'ensemble des participants ; aussi C02 est-il exclu de toutes les analyses et moyennes touchant à l'identification phonémique de ce *continuum*.

Au niveau des groupes et des *continua*, toutes sessions confondues, deux résultats sont à noter : en moyenne, la zone d'ambiguïté est plus grande (équivalent d'une pente de la sigmoïde plus faible) pour le *continuum* /ba-/ma/ que pour /ba-/da/, suggérant ainsi une moins bonne identification du trait phonétique de nasalité que du lieu d'articulation pour les syllabes présentées ici. Aussi, s'il n'y a pas de différence dans les moyennes des deux groupes pour le *continuum* /ba-/da/ (comparaison interne de la partie gauche de la Figure 2), les patients semblent être moins performants pour identifier le /ba-/ma/. Les résultats d'une analyse de variance à mesures répétées viennent renforcer cette observation, par une interaction *groupe* \times *continuum* significative ($F(1,12) = 4.851$; $p = 0.048$). Une hypothèse serait donc que les patients s'améliorent dans l'identification de la nasalité au fur et à mesure des sessions : les aides auditives apporteraient ainsi une compensation par rapport au déficit observé.

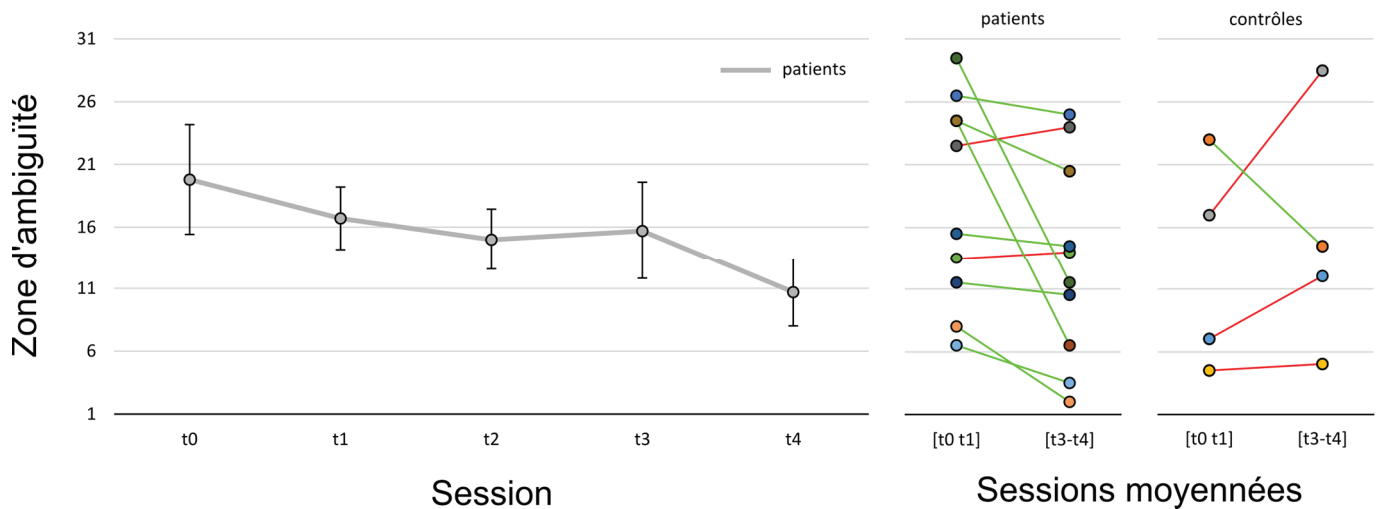


Figure 3. Evolution de la zone d'ambiguïté moyenne du *continuum* /ba-/ma/ pour le groupe *malentendants*, et trajectoire évolutive pour les deux groupes. A gauche, les barres verticales représentent l'erreur-type. A droite, les points permettent de visualiser les trajectoires individuelles entre le début ($t0$ et $t1$) et la fin ($t3$ et $t4$) du protocole, illustrant ainsi le test de Khi-deux (voir ci-dessous). Les connecteurs verts et rouges symbolisent une trajectoire respectivement d'amélioration et de dégradation des performances d'identification phonémique de la nasalité.

Entre la première et la dernière session, la zone d'ambiguïté des patients malentendants pour le *continuum* /ba-/ma/ diminue progressivement (Figure 3, partie gauche). Cet effet longitudinal reste marginal en comparant $t0$ et $t4$ par un test de Wilcoxon ($Z = 1.896$; $p = 0.058$), mais devient significatif en appliquant le même test aux moyennes des deux premières et des deux dernières sessions ($Z = 2.242$; $p = 0.0249$). Afin de contrôler l'absence d'un banal effet d'entraînement qui pourrait expliquer ce résultat, nous avons étudié les trajectoires individuelles entre les deux moyennes de sessions pour nos deux groupes : 8 patients sur 10 améliorent leurs performances, contre 1 sur 4 chez les contrôles. Cette observation ressort au niveau de tendance d'un test de Khi-deux effectué sur ces proportions ($\text{Khi}^2 = 3.764$; $p = 0.052$), suggérant l'absence d'un effet d'entraînement. Les patients semblent donc bien parvenir à mieux identifier les syllabes du *continuum* de nasalité, qui leur posaient particulièrement problème, après deux à quatre mois d'utilisation des aides auditives.

Effets sur la frontière catégorielle.

Un effet de groupe s'observe au niveau de la frontière séparant les catégories phonémiques /ba/, et /da/ ou /ma/, tous *continua* et sessions confondus : les patients ont une frontière plus à gauche pour les deux *continua*, c'est-à-dire qu'ils perçoivent moins de /ba/ que les contrôles (Figure 4). Cet effet du groupe est significatif en opérant

une analyse de variance à mesures répétées avec les facteurs groupe et *continuum* ($F(1,12) = 6.4657$; $p = 0.026$).

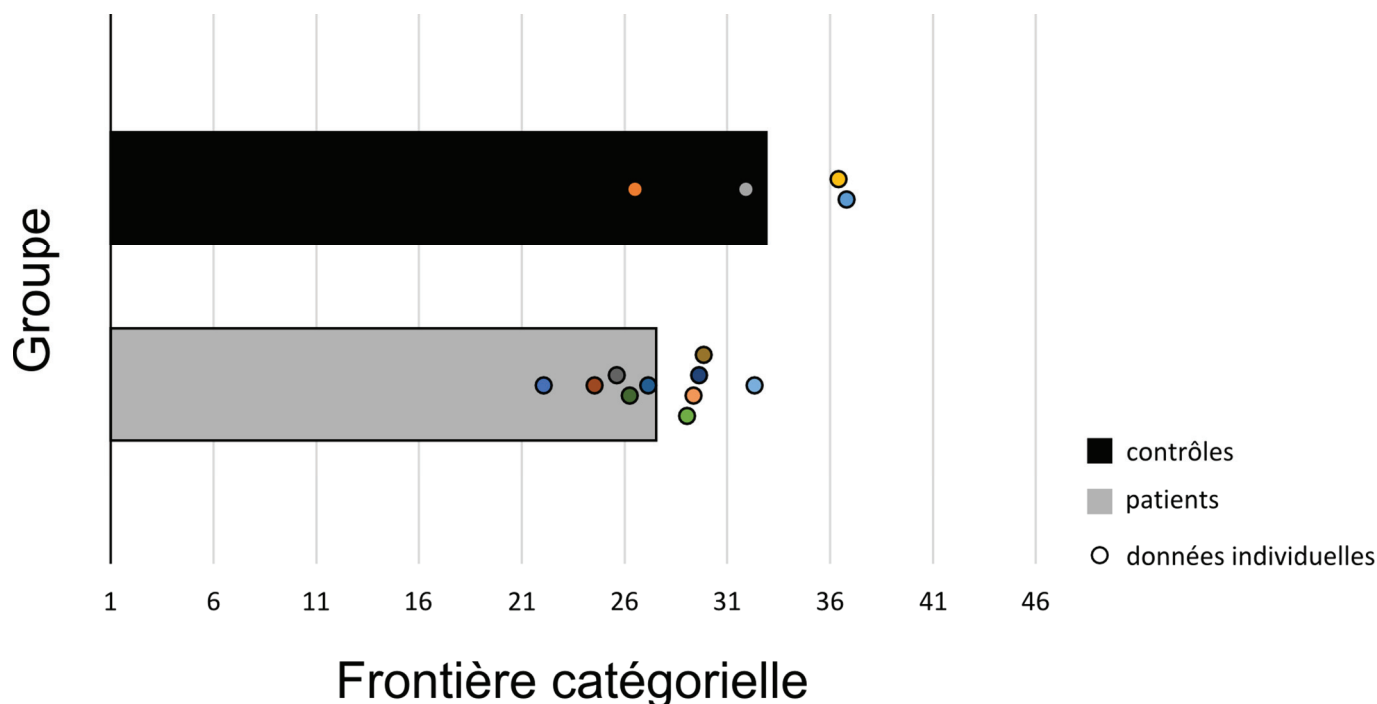


Figure 4. Frontière catégorielle en fonction du groupe, les *continua* et sessions étant confondus (exception faite du participant contrôle C02, dont le point orange correspondant ne couvre que les données du *continuum* /ba/-/ma/). Les points colorés représentent les valeurs individuelles, et la représentation horizontale des barres permet un rapprochement intuitif avec la courbe de perception catégorielle présentée Figure 1.

Au niveau longitudinal, la frontière se décale vers la droite entre $t0$ et $t1$, puis reste stable durant les sessions suivantes. Ainsi, une comparaison entre $t0$ et $t4$ par un Wilcoxon effectué chez les 10 patients fait ressortir cette observation comme une tendance ($Z = 1.944$; $p = 0.052$). Cela ne s'observe que pour /ba/-/ma/ et pour les deux groupes ($\text{Khi}^2 = 2.24$; $p = 0.134$), bien que le décalage $t0-t4$ soit plus grand chez les contrôles. Comme illustré par la Figure 2, les syllabes du *continuum* /ba/-/ma/ sont plus délicates à identifier que celles du *continuum* /ba/-/da/, chez les patients comme chez les contrôles. Par conséquent, nous pouvons imaginer que ce mouvement de frontière est la manifestation d'un changement de référentiel perceptif s'opérant entre $t0$ et $t1$, suite à l'apprentissage de l'indice de nasalité au cours de la tâche d'identification phonémique à $t0$.

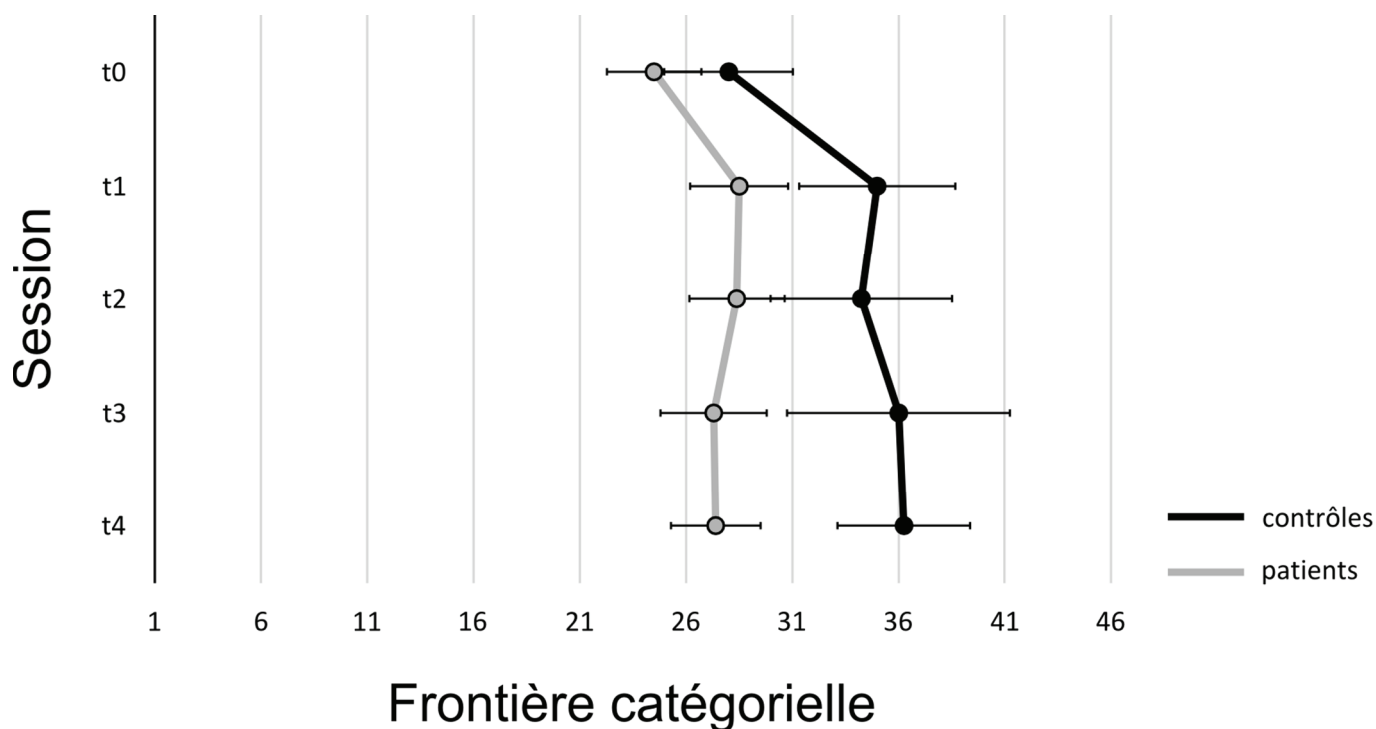


Figure 5. Frontière catégorielle du *continuum* /ba-/ma/ à travers les sessions. Les barres horizontales représentent l'erreur-type.

Par ailleurs, nous retrouvons bien l'effet de groupe dans la Figure 5, avec la frontière des patients décalée à gauche par rapport aux contrôles, dans toutes les sessions. Si un changement de position de la frontière dénote une modification du référentiel perceptif, sa valeur dans l'absolu est plus idiosyncratique que réellement informative quant aux performances d'identification. Une importante variabilité interindividuelle s'observe d'ailleurs sur cette mesure, comme l'illustre bien la Figure 4 avec quelques participants qui ne perçoivent presque que des /ba/ (point taupe à l'extrémité droite des valeurs du groupe contrôle), et d'autres presque uniquement des /ma/ (point marron à l'extrémité gauche des valeurs du groupe *malentendants*).

Effets sur les temps de réponse.

En ajoutant un facteur intra-sujet de position des temps de réponse sur le *continuum* (frontière ou autre ; voir Figure 1 pour une illustration), une analyse de variance à mesures répétées pour la variable de temps de réponse met en évidence un effet significatif dudit facteur : les temps de réponse sont significativement plus lents autour de la frontière, ce qui dénote logiquement la plus grande difficulté à identifier les syllabes ambiguës ($F(1,12) = 6.948$; $p = 0.022$).

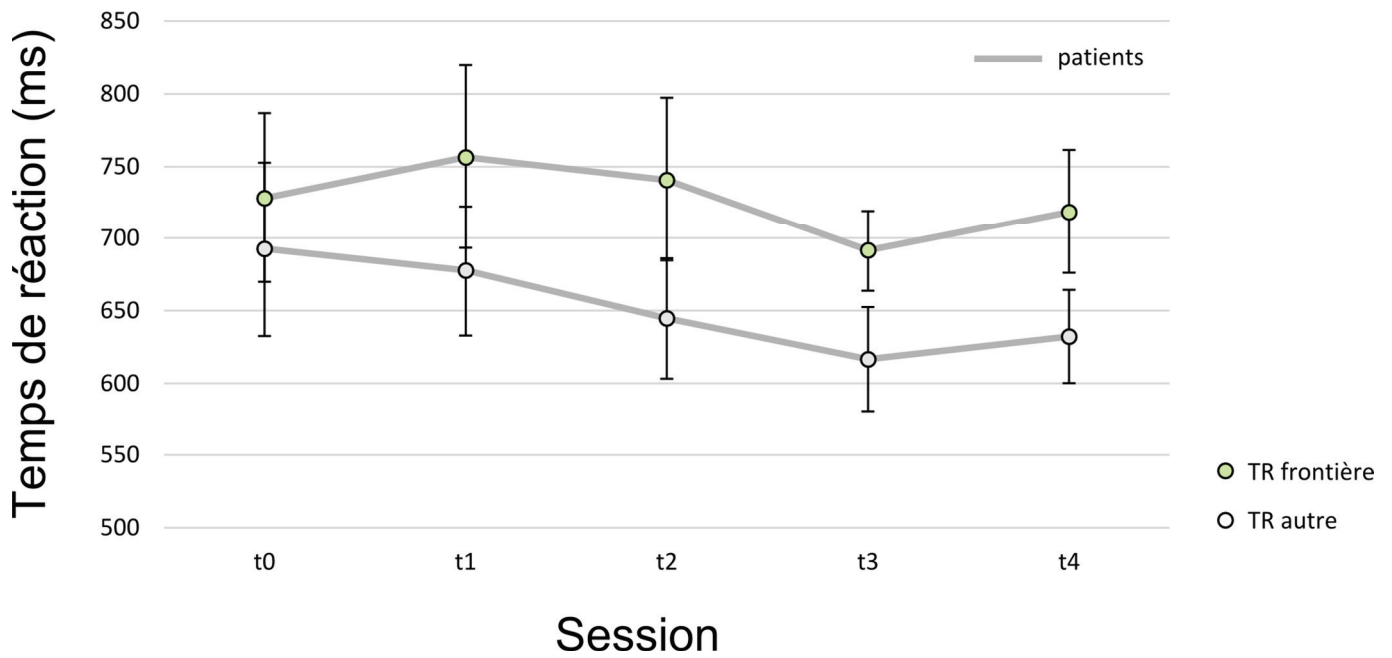


Figure 6. Temps de réaction par rapport à la frontière du *continuum* /ba/-/da/ chez le groupe malentendant et à travers les sessions. Comme dans la Figure 1, les points verts marquent la moyenne des temps de réponse aux 9 syllabes centrées sur la frontière catégorielle, et les points gris la moyenne des temps de réponse aux 41 autres syllabes. Les barres verticales représentent l'erreur-type.

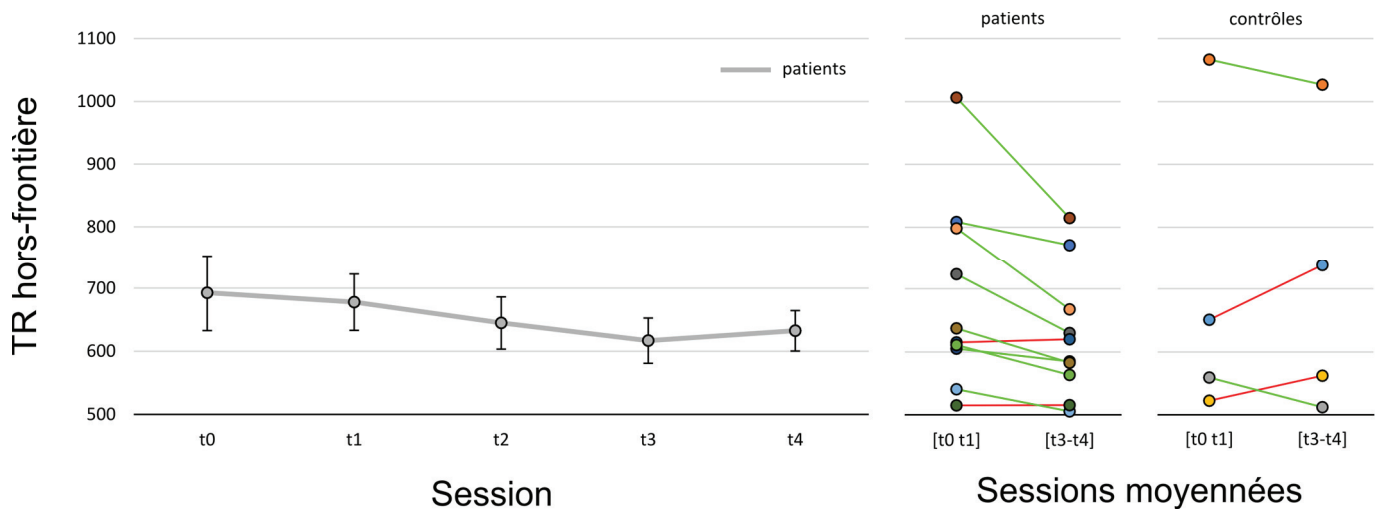


Figure 7. Evolution des temps de réponse moyens pour les syllabes hors de la zone frontière (définie aux 9 stimuli centrés sur la frontière), pour le *continuum* /ba/-/da/ chez les patients malentendants, et trajectoire évolutive pour les deux groupes. Voir la légende de la Figure 3, à la différence que les connecteurs représentent ici le raccourcissement ou l'allongement des temps de réponse aux stimuli non-ambigus. Dans la partie droite, les points orange du contrôle C02 sont présentés à titre strictement illustratif, celui-ci n'étant pas inclus dans les analyses statistiques.

Concernant l'évolution des temps de réponse pour le *continuum* /ba/-/da/ chez les patients, nous observons un raccourcissement spécifique des réponses aux syllabes

éloignées de la frontière (Figure 6), confirmé par les statistiques grâce à l'utilisation d'un test de Wilcoxon comparant $t0$ à $t4$ ($Z = 1.784$; $p = 0.074$) puis, pour confirmer la tendance, la moyenne des deux premières séances à celle des deux dernières ($Z = 2.497$; $p = 0.013$). En analysant les trajectoires individuelles (Figure 7), le test de Khi-deux ne permet toutefois pas de mettre en évidence un effet du groupe ($\text{Khi}^2 = 2.359$; $p = 0.125$).

Le fait que cet effet soit spécifique aux stimuli hors de la zone frontière suggère qu'il ne s'agit pas d'un simple apprentissage moteur, qui influencerait tous les stimuli indépendamment de leur position relative à la frontière, mais plus vraisemblablement d'une identification devenant plus rapide pour les stimuli non-ambigus, donc plus facilement identifiables.

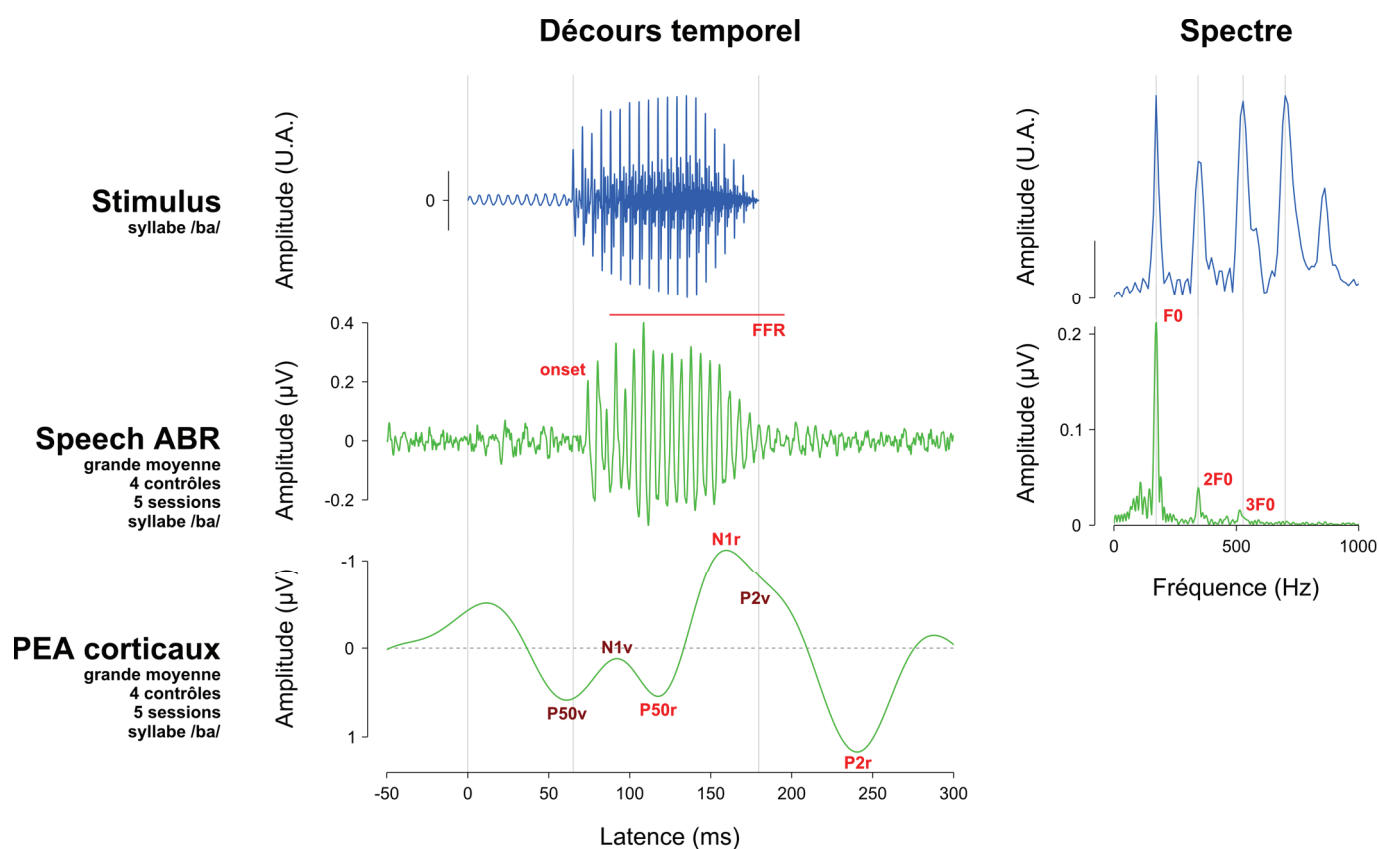
La spécificité de l'effet au *continuum* /ba/-/da/ peut quant à elle trouver une explication dans les moins bonnes capacités d'identification du *continuum* /ba/-/ma/ : les participants, conscients de la plus grande difficulté de la tâche pour le /ba/-/ma/, pourraient prendre davantage de précautions avant de fournir leurs réponses, entraînant ainsi une élévation du temps mesuré pour les stimuli non-ambigus. Une hypothèse alternative pourrait expliquer la spécificité au /ba/-/da/ par la nature de la mesure effectuée : la zone frontière pour le calcul des TR étant fixe (9 stimuli), le *continuum* /ba/-/ma/ contiendra relativement plus de syllabes ambiguës hors de cette zone frontière, augmentant donc mécaniquement la mesure TR hors-frontière. Pour tester cette hypothèse, nous pourrions implémenter une indexation de la taille de la zone frontière sur la pente de la courbe d'identification phonémique, et vérifier si cette nouvelle stratégie de mesure occasionnerait le démasquage de cet effet sur le *continuum* /ba/-/ma/.

Enfin, l'absence de différence significative entre les groupes ne permet pas de conclure quant à un effet exclusivement lié à l'utilisation des aides auditives. Néanmoins, les données individuelles (à droite Figure 7) montrent que la baisse des TR au cours des sessions semble plus systématique chez les patients (8 sur 10) que chez les contrôles (2 sur 4), avec les plus grandes baisses observées chez les patients.

Electrophysiologie.

Description des signaux.

Pour prendre connaissance de la structure de nos différents marqueurs électrophysiologiques, en termes de composantes temporelles et spectrales, nous avons tout d'abord procédé pour chaque groupe au calcul de grandes moyennes à travers les participants, sessions, et syllabes. Notons dès à présent une règle valable pour toutes les figures : nous présentons toujours la Speech ABR en l'électrode Cz (vertex) référencée à la moyenne des mastoïdes, et les PEA corticaux en Cz référencée au nez. Aussi, nous appelons Speech ABR la moyenne des réponses aux stimuli de polarité opposée, désignée dans certains articles comme la FFR d'enveloppe (*envelope FFR*), et qui a pour propriétés de mettre en exergue la partie basse du spectre tout en minimisant l'impact d'un éventuel artefact de stimulation (voir Matériel et méthodes). Enfin, le patient P10 a présenté un artefact du muscle post-auriculaire à toutes les sessions d'enregistrement ; cet artefact ne s'observe que sur la bande de fréquence de la Speech ABR (entre 80 et 1000 Hz), qu'il détériore fortement du fait de sa localisation aux électrodes mastoïdiennes (indispensables au re-référencement du signal) et de son amplitude élevée. Nous avons donc exclu P10 de toutes les moyennes et analyses concernant la Speech ABR.



recueil, moins les rejets) ; pour les PEA corticaux, elle en est composée de 2852 (200 essais par recueil pour un total théorique de 4000 ; plus d'essais sont cependant rejetés en proportion que pour les Speech ABR, à cause des artefacts oculaires impactant particulièrement la bande de fréquence des PEA corticaux, ici 3 à 20 Hz). Les traits verticaux gris permettent de faciliter la comparaison entre les différents signaux ; aussi les appellerons-nous des *repères*. Les PEA corticaux sont représentés avec la négativité vers le haut, par convention.

Sur la première ligne de la Figure 8, nous observons le voisement de la syllabe *ba* entre le premier et le deuxième repère, durant environ 60 ms, suivi d'un saut d'amplitude à 65 ms (deuxième repère) traduisant le relâchement de l'occlusion bilabiale (*release*). La voyelle s'étend ensuite jusqu'à 180 ms (troisième repère). La Speech ABR reproduit bien ce schéma, avec une composante transitoire de type *onset response* à 74,2 ms, soit 9 ms après l'explosion correspondante dans le stimulus, suivie d'une composante soutenue, la FFR, qui reprend la fréquence fondamentale de 170 Hz. Cette fréquence prédomine effectivement sur le spectre, qui révèle aussi les deux premières harmoniques (340 et 510 Hz). Une activité évoquée par le voisement semble enfin présente, bien que moins structurée que les deux composantes *onset* et FFR.

Les PEA corticaux présentent deux séries d'ondes P50-N1-P2, en réponse au début du voisement (labels des ondes avec *v* pour voisement, en bordeaux Figure 8) et à l'explosion de la consonne à 65 ms (labels avec *r* pour *release* ; en rouge Figure 8). La P2 au voisement, censée arriver aux alentours de 200 ms, est probablement la cause de l'épaule observée sur le front de descente de la N1 à l'explosion (troisième repère).

Effets du groupe.

La réponse moyenne à tous les stimuli permet de mettre en évidence les différences majeures entre les deux groupes. Pour la Speech ABR, nous observons chez les patients une réponse *onset* réduite en amplitude et décalée en latence (Figure 9, ligne du milieu, à gauche) ; si l'amplitude anormale traduit probablement un déficit dans l'encodage des hautes fréquences, nécessaires pour l'émergence rapide de pics, l'observation sur la latence ne s'explique pas aussi aisément : il peut s'agir du simple effet de moyenne d'un plus grand nombre de sujets, d'une déficience de l'encodage neuronal temporel, ou de l'effet de la carence en hautes fréquences, entraînant la synchronisation d'un effectif plus faible de neurones dans les voies auditives ascendantes et en conséquence une période entière de retard par rapport aux contrôles. La même observation peut être faite à la conclusion de la FFR, celle des patients disparaissant plus tôt. Dans le spectre de la FFR, l'amplitude du 170 Hz est plus élevée chez les contrôles. En y regardant de plus près, le niveau de bruit dans cette bande de fréquence est également plus élevé chez les contrôles.

De fait, une normalisation par rapport de puissances (RMS du pic à 170 Hz divisée par RMS du bruit alentour) renverse l'observation, avec une surreprésentation du 170 Hz chez les patients (Figure 10).

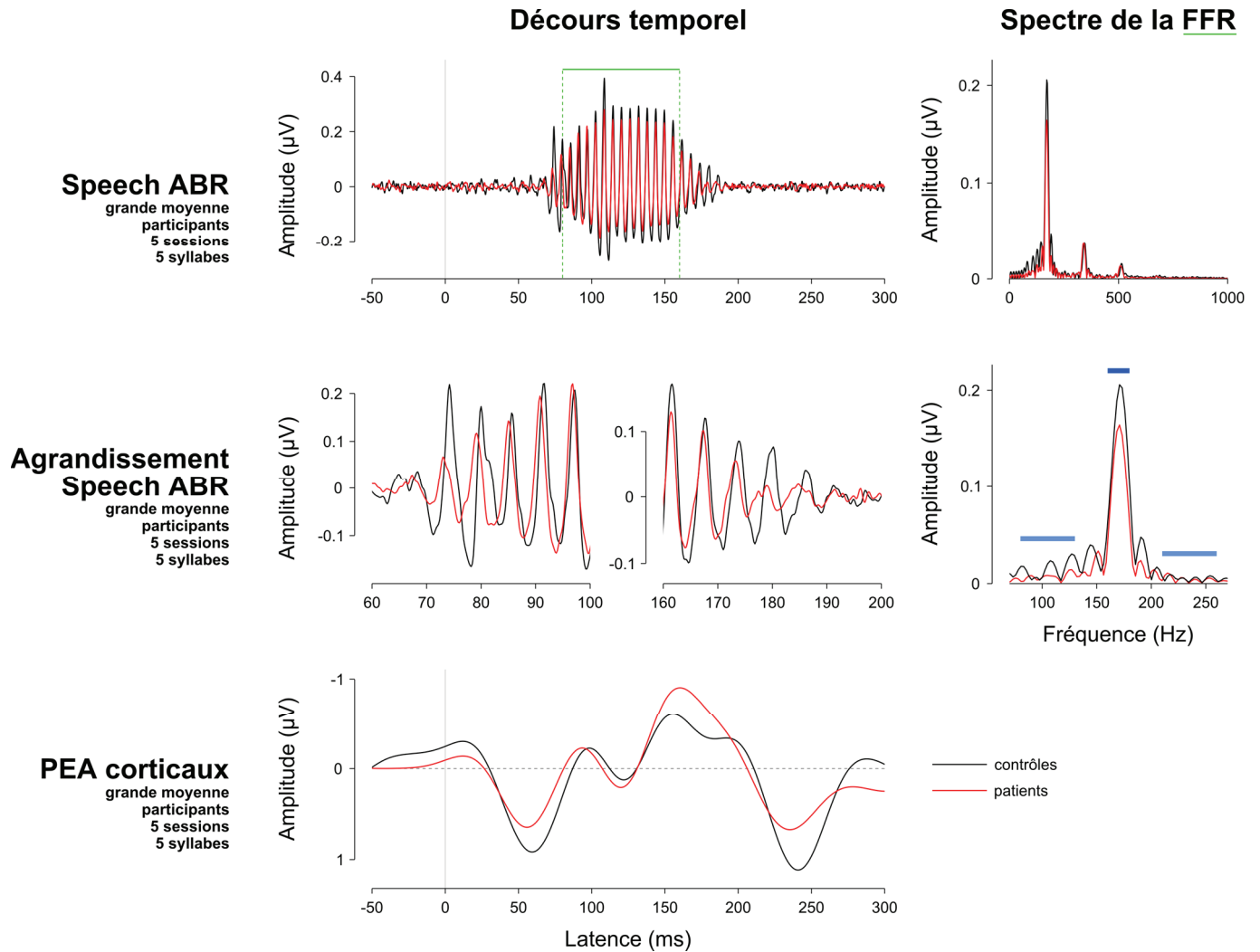


Figure 9. Grandes moyennes des Speech ABR (lignes du haut et du milieu) et PEA corticaux (ligne du bas) à travers sessions et syllabes pour chaque groupe. Les réponses des patients sont présentées en rouge. La deuxième ligne reprend les données de la première, en réduisant la fenêtre temporelle (à gauche) ou spectrale (à droite) d'observation. La Speech ABR moyenne des contrôles est composée de 190140 essais, celle des patients de 419369. Les PEA corticaux regroupent respectivement 14467 et 35032 essais pour les contrôles et les patients. Le spectre de la FFR (colonne de droite) est calculé dans l'intervalle temporel entre 80 et 160 ms, marqué en vert sur les Speech ABR (ligne du haut). Dans l'agrandissement de ce spectre (ligne du milieu, à droite), le trait bleu foncé matérialise la fenêtre de calcul de la puissance moyenne (*Root Mean Square*) du pic à 170 Hz, et les traits bleus la fenêtre de bruit, afin de normaliser le pic f_0 (voir ci-dessous).

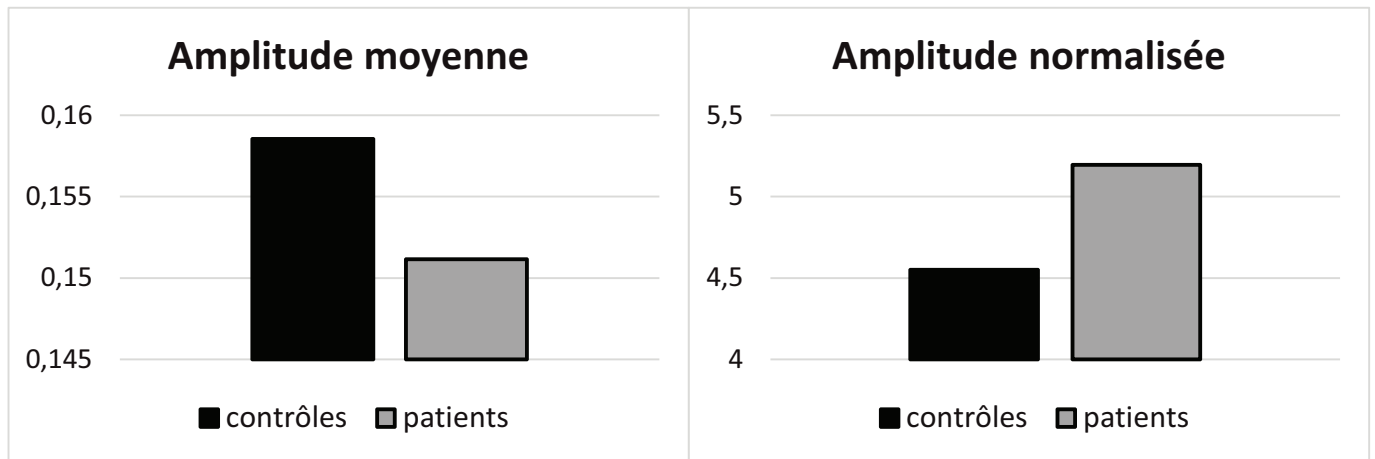


Figure 10. Amplitude du 170 Hz dans le spectre de la FFR, en fonction du groupe et de la mesure. L'amplitude moyenne prend la mesure sur une fenêtre de 160 à 180 Hz, tandis que l'amplitude normalisée consiste en un rapport de puissances (voir Figure 9). A noter que les valeurs présentées ici sont la moyenne de mesures effectuées sur chaque Speech ABR de chaque participant, stimulus et session ; il ne peut donc y avoir d'effet de déphasage entre les FFR des différents participants, qui entrainerait une baisse de l'amplitude de cette réponse dans la grande moyenne.

La prise en compte du bruit permet certainement de compenser pour la différence d'effectif (4 contrôles contre 10 malentendants), la grande moyenne d'un plus grand nombre de réponses permettant relativement mieux de se débarrasser du bruit, temporellement inconstant. Nous reproduisons donc ici grâce aux données normalisées les résultats d'une étude antérieure ayant mis en évidence un encodage des basses fréquences plus important chez des patients atteints de surdité neurosensorielle, dans une tranche d'âge similaire (60 à 71 ans ; Anderson *et al.*, 2013).

Au niveau des PEA corticaux, des différences entre les groupes sont visibles au niveau de P50v (P50 au voisement), N1r (N1 au relâchement de l'occlusion), P2r et éventuellement P2v, confondu à N1r (Figure 9). Pour P50v et P2r, l'amplitude est plus faible chez les patients, et la latence raccourcie de quelques millisecondes. La N1r est quant à elle plus ample et de morphologie plus simple que chez les contrôles, notamment sans le rebond positif imputable à P2v. L'amplitude et la latence de ces différentes ondes seront donc à analyser à travers les sessions pour mettre en évidence un éventuel effet de plasticité ; néanmoins avec la difficulté d'une très importante variabilité interindividuelle dans la structure des PEA corticaux : certains participants montrent bien toutes les ondes, et d'autres n'ont que le complexe P50-N1-P2 en réponse à l'explosion de la consonne.

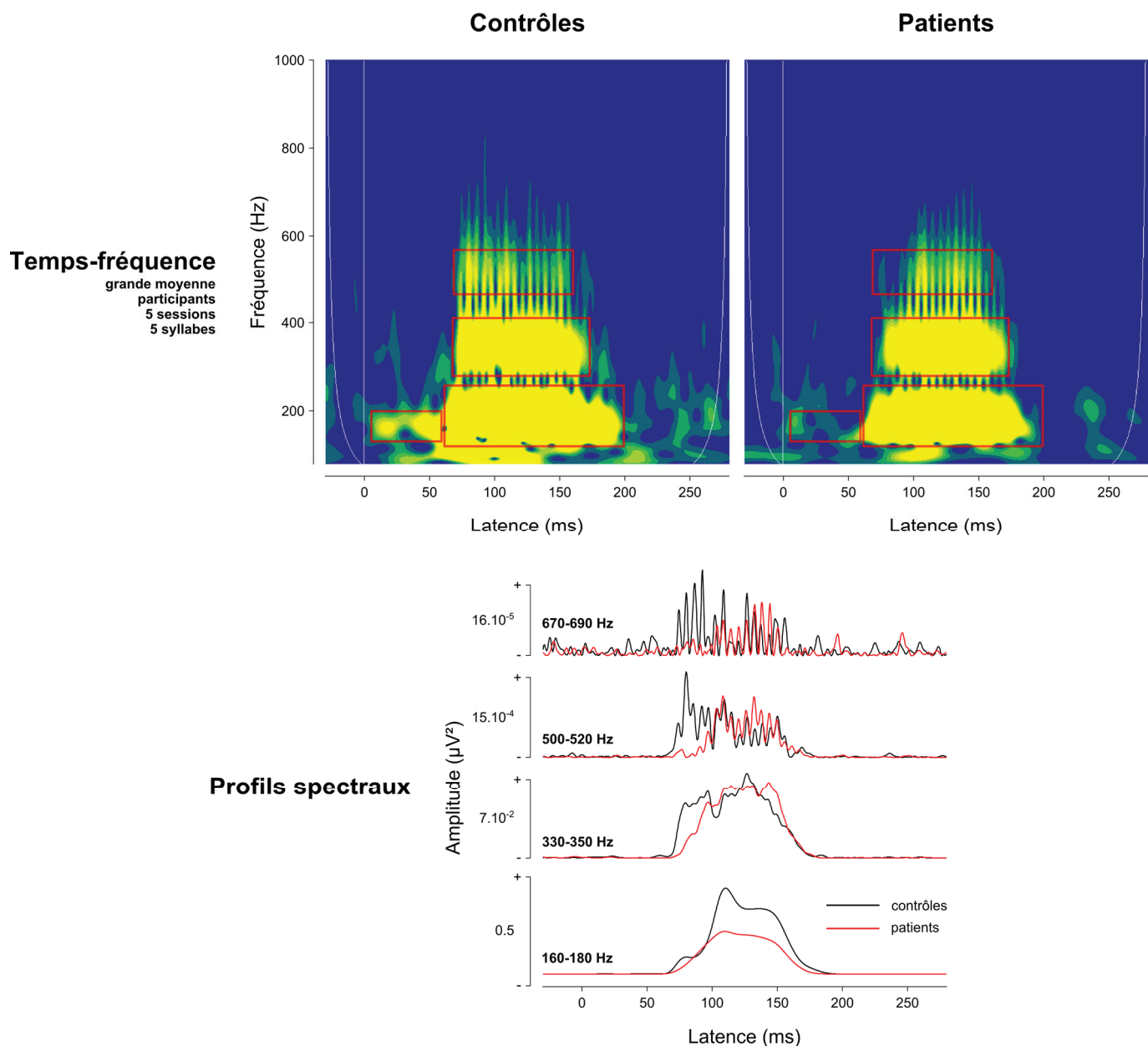


Figure 11. Grandes moyennes des Speech ABR à travers sessions et syllabes, observées en temps-fréquence (transformée en ondelettes) dans les deux groupes de participants (échelles identiques). Les cadres rouges font office de repères pour simplifier la comparaison entre les groupes. A droite, le dérours temporel est analysé pour des bandes de fréquence correspondant à F0 et ses trois premières harmoniques.

L'étude des Speech ABR de nos deux groupes en temps-fréquence apporte des éléments complémentaires par rapport aux observations précédentes. Chez les contrôles, une activité FFR existe pendant le voisement (Figure 11, partie haute, cadre rouge de gauche), cependant bien moins ample que la FFR à la voyelle ; ceci est à rapprocher de l'activité observée Figure 8 en réponse au *ba*, avec prudence puisqu'ici les syllabes sont moyennées : cette FFR pourrait être apportée par la réponse au murmure nasal de *ma30* et *ma70*, plus ample que le voisement des consonnes /b/ et /d/ (voir ci-dessous). Notons

qu'elle ne se trouve pas chez les patients, malgré un plus grand nombre d'essais moyennés. Cela pourrait être lié aux dommages cochléaires liés à la presbyacousie, qui entraînent une baisse de la quantité de fibres du nerf auditif recrutés à l'écoute d'un son à une intensité donnée, et par conséquent l'élévation du seuil d'intensité pour le déclenchement de l'activité FFR.

Au niveau de l'*onset*, le front de montée est raide et net chez les contrôles, avec au moins quatre composantes spectrales démarrant de concert (F0, 2F0, 3F0 et 4F0 ; cadres rouges superposés sur les temps-fréquences et profils spectraux pour F0, 2F0 et 3F0 ; profil spectral seul pour 4F0) ; chez les patients, leur apparition est plus progressive, l'activité soutenue apparaissant séquentiellement sur F0, 2F0 puis 3F0 (Figure 11, partie basse). Cela illustre l'observation faite sur les décours temporels, Figure 8, et est compatible avec l'hypothèse d'une activité de suivi de phase (phase-locking) plus lente à mettre en route chez le malentendant, avec une intensité acoustique requise plus élevée pour le recrutement d'un effectif suffisant de neurones dans les voies auditives ascendantes.

L'extinction anticipée de la FFR chez les patients trouve elle aussi un nouvel éclairage avec les temps-fréquences. L'activité soutenue s'éteint d'abord à 3F0, puis à 2F0, aux mêmes latences dans les deux groupes. C'est au niveau du 170 Hz qu'une différence s'observe : la FFR y persiste plus longtemps chez les contrôles (côté droit du plus grand cadre rouge), en réponse à des périodes de faible amplitude (voir la fin du stimulus en Figure 8). L'hypothèse du seuil d'intensité est compatible avec cette observation : les dernières périodes du stimulus seraient ainsi trop faibles pour compenser les carences en cellules ciliées fonctionnelles dans la cochlée des patients malentendants.

Effets de la syllabe et du groupe.

Jusqu'à présent, nous n'avons observé l'activité électrophysiologique qu'en réponse au *ba* ou aux syllabes mélangées dans une grande moyenne. Nous allons maintenant nous intéresser aux différences d'encodage neuronal existant entre les 5 syllabes *ba*, *da30*, *da70*, *ma30* et *ma70*. Rappelons que si *ba* est physiquement identique chez tous les participants, les quatre autres syllabes sont propres à chacun d'eux, étant sélectionnées sur la base des résultats de la tâche d'identification phonémique. Dans cette première vague d'analyse, nous n'avons pas pris cette variation en compte, qui pourrait cependant expliquer certaines observations : comme nous l'avons vu dans la partie comportementale, les patients perçoivent moins de /ba/ que les contrôles, et ont plus de mal à identifier le trait de nasalité ; ils auront donc une sélection préférentielle de certains stimuli par rapport aux contrôles, ce qui pourrait renforcer certaines composantes temporelles ou

spectrales dans les grandes moyennes des réponses électrophysiologiques à l'étude dans cette présente partie.

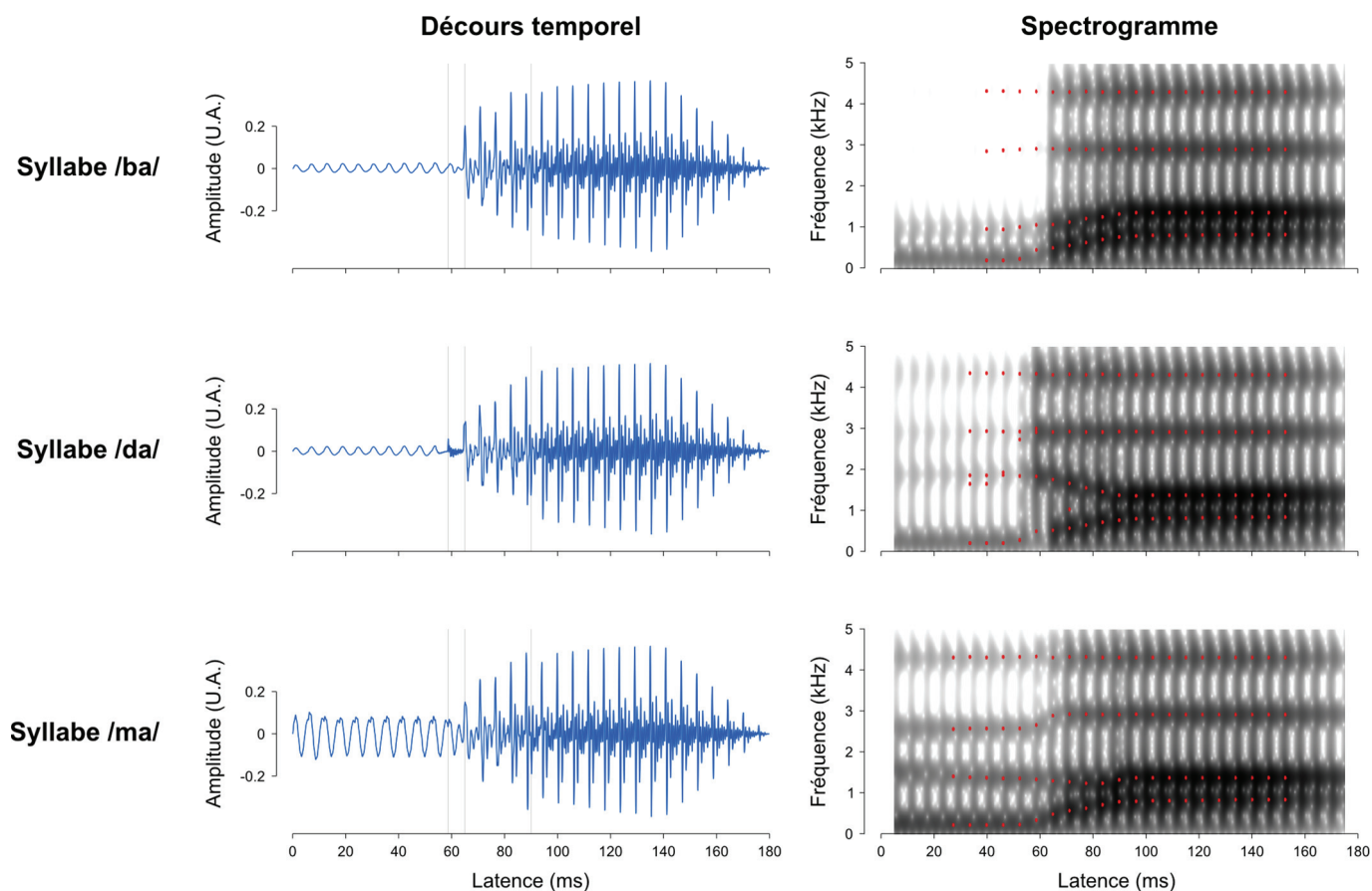


Figure 12. Description des syllabes /ba/, /da/ et /ma/ aux extrémités des *continua*. Des repères sont ajoutés (lignes verticales grises) pour faciliter la comparaison entre les signaux.

Pour comprendre les particularités des réponses électrophysiologiques aux différentes syllabes de synthèse, intéressons-nous d'abord à la structure de celles-ci en comparaison avec le /ba/ que nous avons décrit à l'occasion de la Figure 8. L'explosion du /da/ a ainsi lieu un cycle plus tôt (premier repère Figure 12), et marque le début d'une trajectoire descendante du deuxième formant, là où une trajectoire ascendante entraîne préférentiellement la perception du phonème /b/. Le /ma/ possède un murmure nasal plus ample que le voisement (oral) de syllabes /ba/ et /da/, et spectralement plus riche ; aussi, la désocclusion apparaît relativement moins contrastée par rapport au /ba/, malgré une amplitude absolue proche (deuxième repère). Enfin, la voyelle /a/ de ces trois syllabes est physiquement identique à partir de 90 ms (troisième repère).

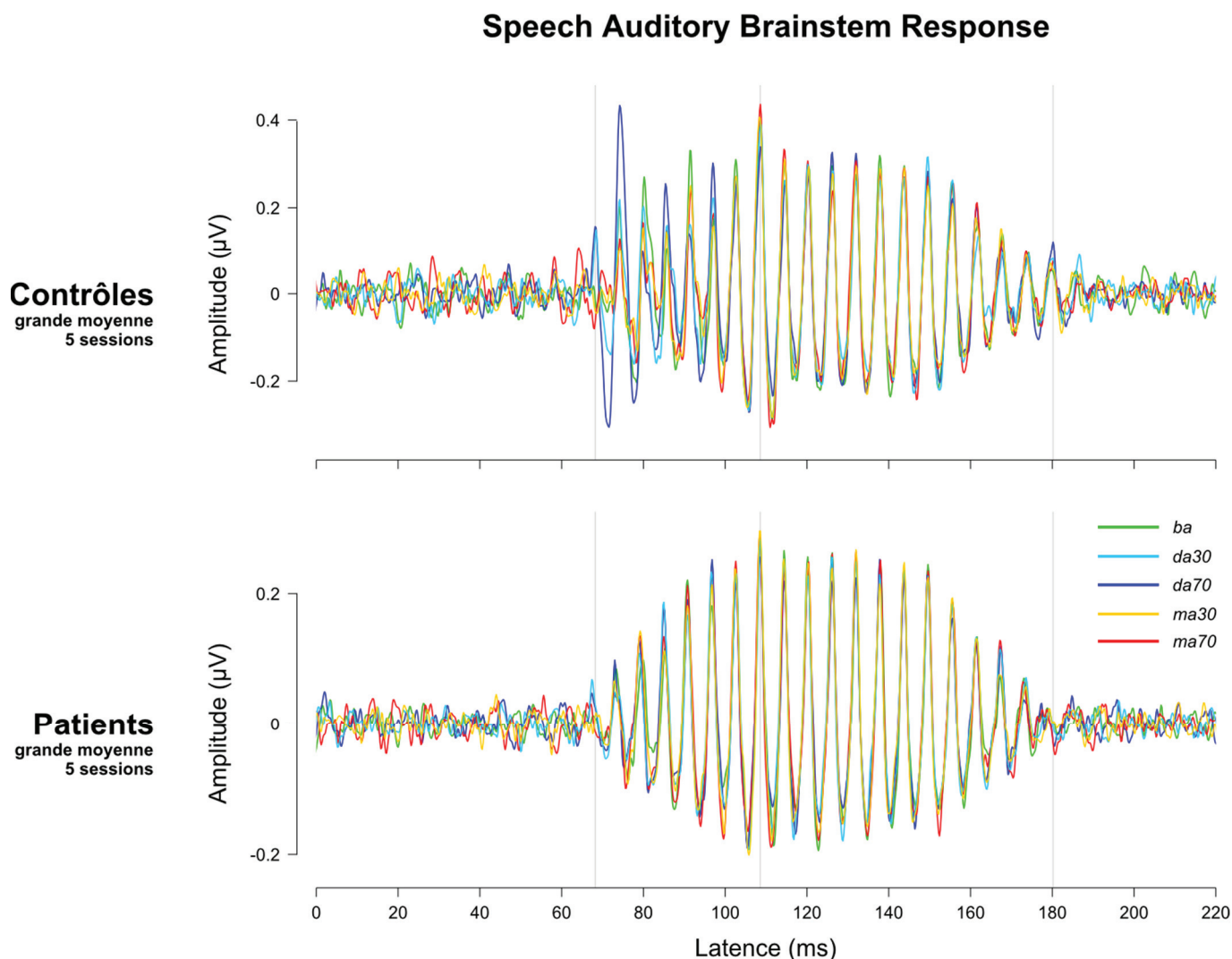


Figure 13a. Grande moyenne des Speech ABR évoquées par les 5 syllabes pour chacun des deux groupes, toutes sessions confondues. Des repères sont ajoutés (lignes verticales grises) pour faciliter la comparaison entre les signaux. Les contrôles ont en moyenne 38028 essais par Speech ABR, et les patients environ 83874. Notons que pour favoriser la visibilité des réponses, les échelles d’amplitude sont différentes entre contrôles et patients.

Trois observations s’imposent à nous dans cette représentation des Speech ABR aux différentes syllabes (Figure 13a). La première se situe au niveau de l’*onset* : le *da70* se démarque très clairement des autres syllabes chez les contrôles (voir les deux pics négatif et positif immédiatement après le premier repère), suivi du *da30* avec une amplitude semblable au *ba* à l’endroit du premier repère, et une amplitude amoindrie aux deux pics suivants. Ce fort trait distinctif n’apparaît aucunement chez les patients, sans doute car faisant appel aux hautes fréquences (voir le front de montée des formants 3 et 4, assimilable à un clic, dans le spectrogramme du /da/ en Figure 12). La deuxième observation est le caractère hautement superposable de la FFR des Speech ABR aux

différentes syllabes, particulièrement après 108 ms environ (deuxième repère), soit 18 ms après le début de la phase de la voyelle /a/ identique entre les syllabes. Enfin, la troisième observation est à nouveau l'arrêt précoce de la FFR chez les patients, montré en réponse à toutes les syllabes (troisième repère).

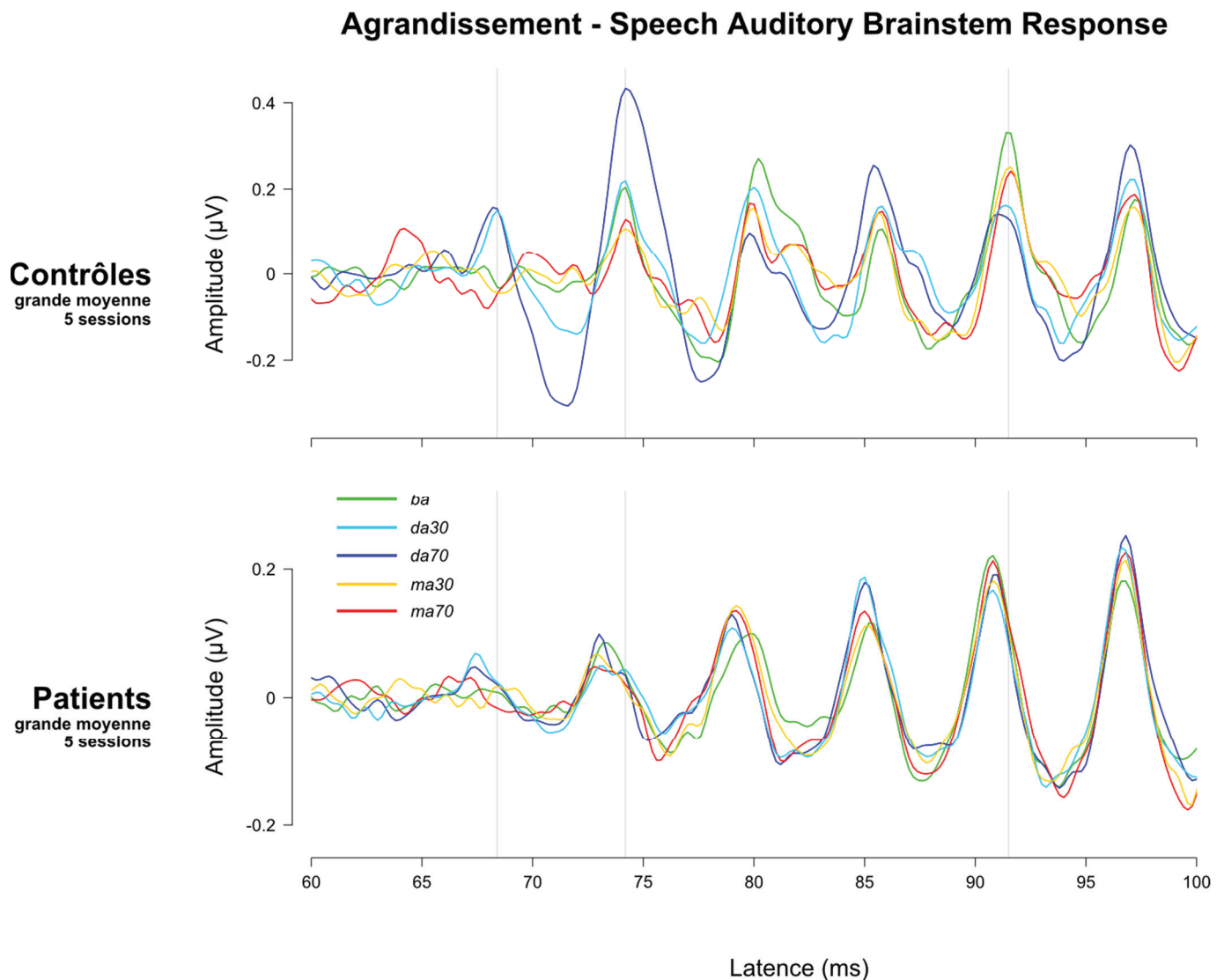


Figure 13b. Agrandissement des Speech ABR de la Figure 13a. Le premier repère correspond également au premier repère dans la Figure 13a. Les échelles d'amplitudes sont identiques à celles utilisées Figure 13a, et différent donc entre contrôles et patients.

En regardant de plus près la zone de signal près de l'*onset* (Figure 13b), il apparaît que les pics évoqués par les syllabes similaires (*da30* et *da70*, ou *ma30* et *ma70*) ont tendance à être proches, comme c'est le cas pour les /da/ au niveau du premier repère, autour de 85 ms et au troisième repère, et pour les /ma/ au niveau des deuxième et troisième repères, chez les participants du groupe contrôle. Nous pouvons également

observer une monotonie (*ba*, puis *da30*, puis *da70*, par exemple) à certains endroits, tels qu'entre les deux premiers repères, ou à l'emplacement du troisième. Chez les patients, les syllabes sont beaucoup plus difficilement distinguables : juste avant le premier repère, les deux /da/ se démarquent ; aussi, au niveau du pic à 85 ms, les syllabes sont identifiables sur la base de leur pic. Mais ces observations semblent bien moins robustes que le large pic des contrôles pour le *da70* au niveau du deuxième repère, complètement absent chez les patients comme nous l'avons vu dans la Figure 13a.

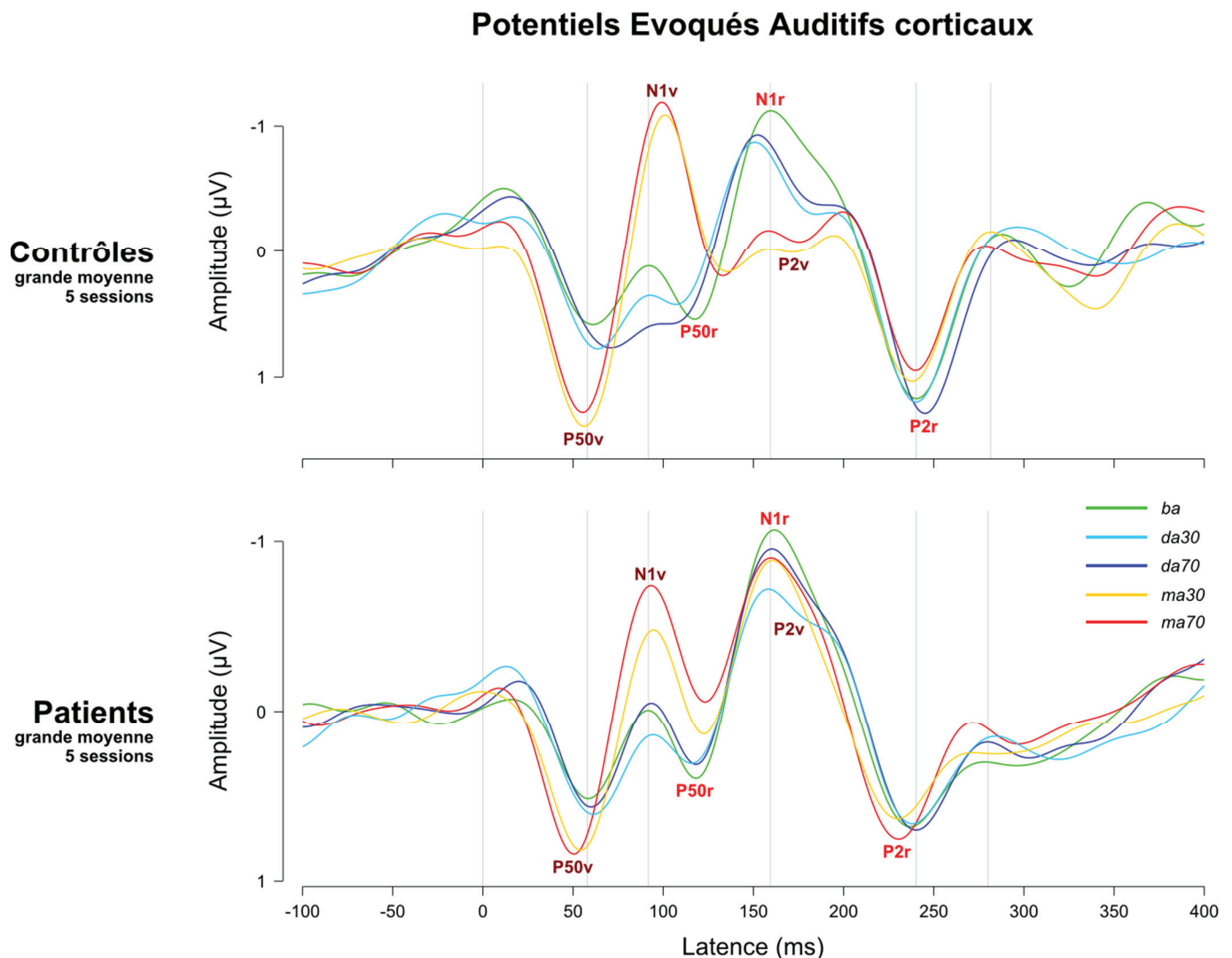


Figure 14. Grandes moyennes des PEA corticaux en fonction des stimuli et du groupe, toutes sessions confondues. Les contrôles ont en moyenne 2893 essais par courbe, et les patients 7006. Les échelles d'amplitude sont différentes pour les deux groupes, afin de rendre l'observation des PEA corticaux des patients plus aisée (même logique que pour les Figures 13a et 13b).

Globalement, la structure des PEA corticaux enregistrés chez les patients est plus simple, en double W (Figure 14, en bas). La P50r y est plus facilement observable que

chez les contrôles (Figure 14, en haut), mais en partie à cause d'une latence commune entre les différentes syllabes, ce qui peut être considéré comme un manque d'information. Les contrôles montrent une P2v au voisement, particulièrement pour les /ma/ qui ont le murmure nasal plus ample : l'absence d'une réponse comparable chez les patients peut être le signe d'une déficience perceptive des hautes fréquences caractéristiques de ce phénomène phonétique. En effet, sans hautes fréquences, le murmure nasal ressemble au voisement des syllabes /ba/ et /da/ (voir les spectrogrammes Figure 12). Concernant les amplitudes, le complexe P50v-N1v des patients est moins grand, et l'onde P2r émerge moins. Par contre, la N1r a tendance à être de même amplitude, voire même d'amplitude supérieure à son équivalent chez les contrôles ; cela peut provenir d'un impact moindre de P2v, normalement superposée au front de descente de la N1r, du fait de la présence restreinte des hautes fréquences chez le malentendant.

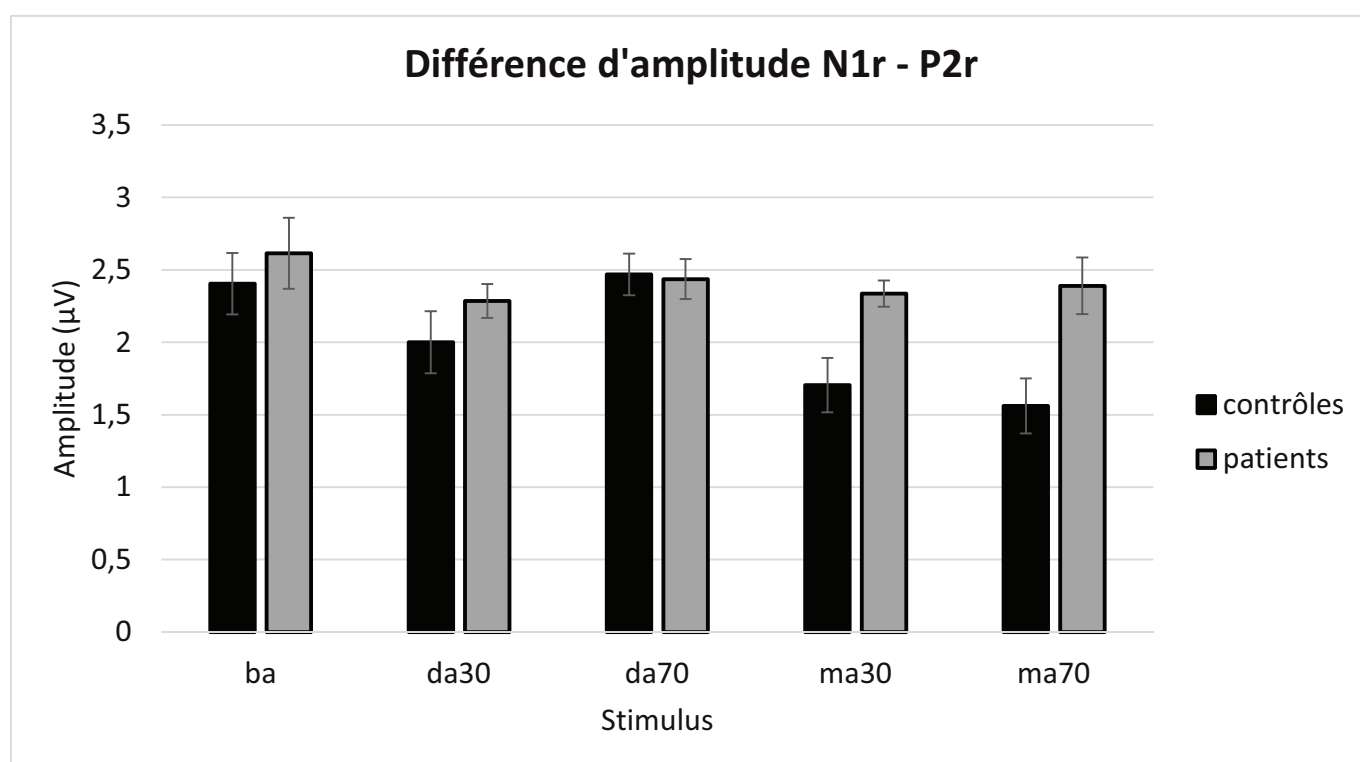


Figure 15. Différence d'amplitude N1r-P2r en fonction du stimulus et du groupe. Les patients sont représentés en gris. Les barres d'erreur indiquent l'erreur-type.

Aussi, la différence d'amplitude entre N1r et P2r montre de plus grandes valeurs spécifiquement aux deux syllabes /ma/ chez les patients (Figure 15), ce qui provient sans doute de l'absence d'épaule sur la N1r évoquée par *ma30* et *ma70* (voir Figure 14).

Analyse des effets longitudinaux.

Speech ABR : évolution du contenu spectral de la FFR (moyenne des 5 stimuli).

Deux mesures ont été relevées sur la Speech ABR de chaque participant à l'exception de P10, pour chaque stimulus, et à chaque session : le rapport signal-sur-bruit (SNR), et l'amplitude normalisée du pic à 170 Hz dans le spectre de la FFR ainsi que des pics à chacune des trois premières harmoniques (voir Figure 9).

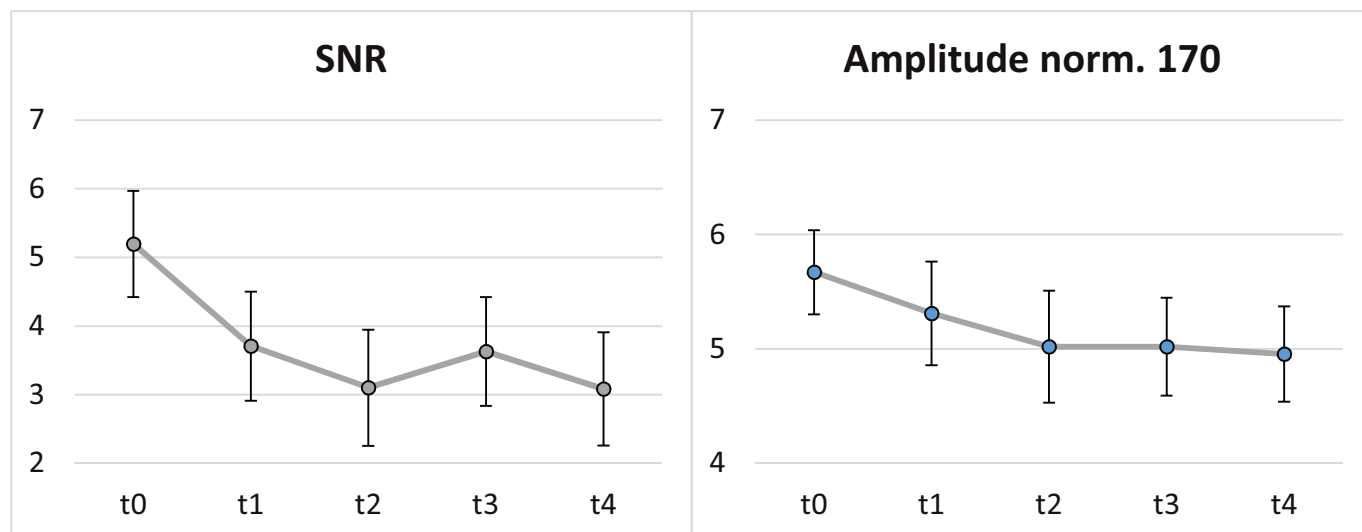


Figure 16. Evolution du SNR et de l'amplitude normalisée du 170 Hz chez les patients, tous stimuli confondus. Rappelons que du fait de la présence d'un artefact musculaire de très grande amplitude dans toutes ses Speech ABR, le patient P10 est exclu des analyses concernant cette réponse.

A travers les quatre mois du protocole, le SNR des Speech ABR diminue significativement chez les patients, tous stimuli confondus (test de Wilcoxon comparant les valeurs à t_0 et t_4 ; $Z = 2.192$; $p = 0.028$; Figure 16, à gauche). Or la fenêtre *signal + bruit* du SNR est mesurée au niveau de la FFR, dont la fréquence prépondérante est F_0 (voir Figure 11). Pour tester l'hypothèse que cette baisse du SNR serait due à une diminution de l'encodage de F_0 , nous avons suivi l'amplitude normalisée de cette fréquence au cours des sessions : l'encodage de F_0 dans la FFR présente effectivement une tendance à la diminution chez les patients appareillés (test de Wilcoxon entre t_0 et t_4 pour 9 malentendants ; $Z = 1.836$; $p = 0.066$; Figure 16, à droite). Cet effet n'étant pas significatif pour les autres composantes spectrales ($p > 0.44$), ce résultat va donc bien dans le sens d'une compensation de la surreprésentation des basses fréquences observée en comparaison aux participants contrôles. Cependant, le test de Khi-deux ne permet pas de différencier les groupes (non montré), les contrôles montrant aussi une diminution de ces deux mesures entre t_0 et t_4 , malgré une évolution irrégulière en considérant l'ensemble des sessions.

PEA corticaux : évolution du complexe N1r-P2r pour chacun des cinq stimuli.

Dans les réponses de chaque participant, pour chaque stimulus et à chaque session, nous avons mesuré par marquage manuel la latence et l'amplitude des ondes N1r et P2r qui s'observaient sans ambiguïté chez tous les participants.

L'utilisation d'ANOVA à mesures répétées nous permet d'estimer la part des facteurs groupe (facteur inter-sujets), stimulus, et session (facteurs intra-sujet) dans la variance des deux paramètres décrivant ces réponses. Nous rapportons ici uniquement les effets et interactions significatifs.

Un premier effet s'observe sur la latence de l'onde N1r avec une interaction *stimulus* \times *groupe* significative ($F(4,48) = 3.099$; $p = 0.024$). Un test post-hoc (Fisher LSD) révèle que la différence entre les groupes est significative pour les syllabes *ma30* et *ma70* ($p = 0.03$ dans les deux cas), les patients présentant des latences d'environ 50 ms plus précoces pour celles-ci, probablement du fait d'une onde P2v moins ample comme vu précédemment (Figure 14).

L'amplitude de l'onde N1r a tendance à être plus élevée chez les patients ($F(1,12) = 3.83$; $p = 0.074$), et varie significativement en fonction du stimulus ($F(4,48) = 7.683$; $p < 0.001$) avec *da30*, *ma30* et *ma70* entraînant une onde de plus faible amplitude ($p < 0.02$). Comme nous pouvons le constater dans la Figure 14, il s'agit bien des stimuli entraînant les N1r les moins amples, d'autant plus pour les deux /ma/ chez les contrôles.

La latence de l'onde P2r ne montre pas d'effet (ni d'interaction) des facteurs groupe, stimulus ou session. Par contre, l'ANOVA sur l'amplitude de la P2r révèle une interaction *stimulus* \times *session* \times *groupe* significative ($F(16,192) = 2.359$; $p = 0.003$). Puis, en restreignant l'ANOVA au groupe de patients, une interaction marginale *stimulus* \times *session* apparaît ($F(16,144) = 1.551$; $p = 0.09$).

Cette dernière interaction devient significative en appliquant l'ANOVA non plus à l'amplitude de l'onde P2r, mais à l'amplitude différentielle N1r-P2r ($F(16,144) = 1.87$; $p = 0.028$) ; l'investigation des tracés évolutifs suggère un effet mené par le stimulus *ma70*, dont l'amplitude différentielle N1r-P2r diminue régulièrement au cours des sessions (Figure 17, courbe rouge). Ce résultat va donc dans le sens d'une compensation de l'écart d'amplitude anormalement élevée entre N1r et P2r à l'explosion du /ma/, chez les patients par rapport aux contrôles. Des tests post-hoc (Fisher LSD) confirment cette interprétation, avec une différence significative entre *t0* et *t4* pour le stimulus *ma70* ($p = 0.01$), marginale pour *da30* ($p = 0.06$), mais non-significative pour les trois autres stimuli ($p > 0.18$).

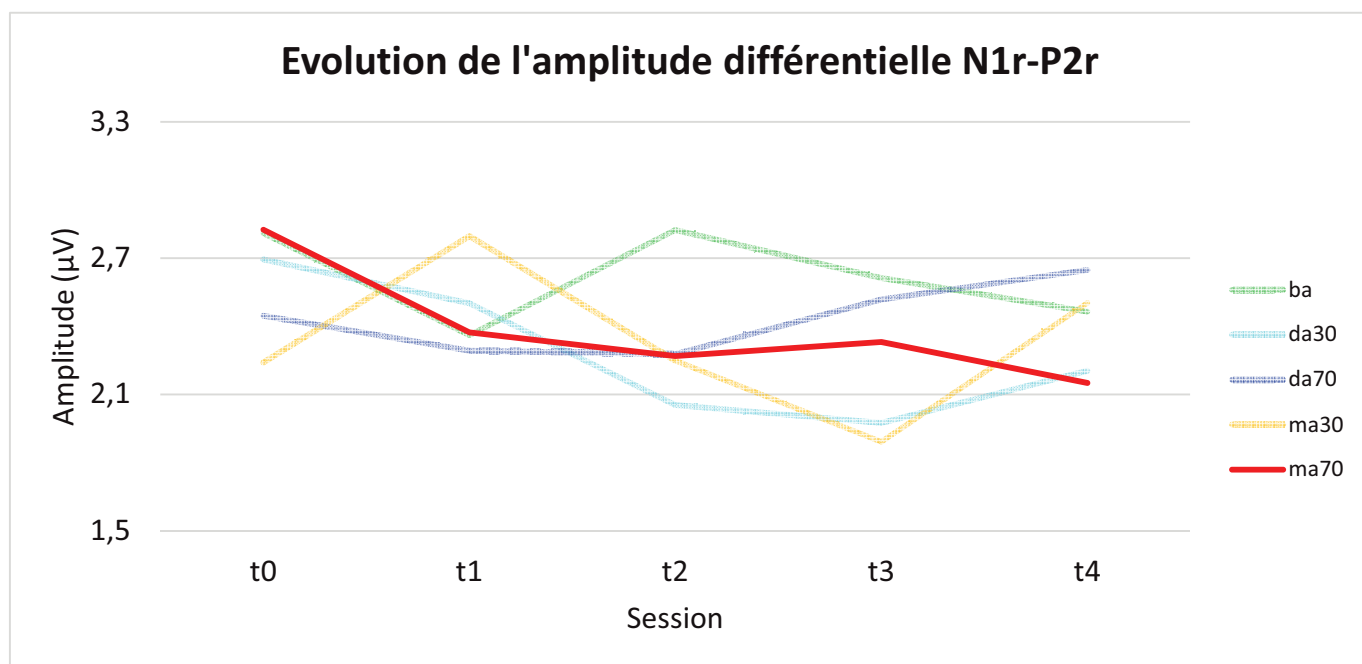


Figure 17. Evolution du différentiel d'amplitude N1r-P2r chez les 10 patients en fonction du stimulus.

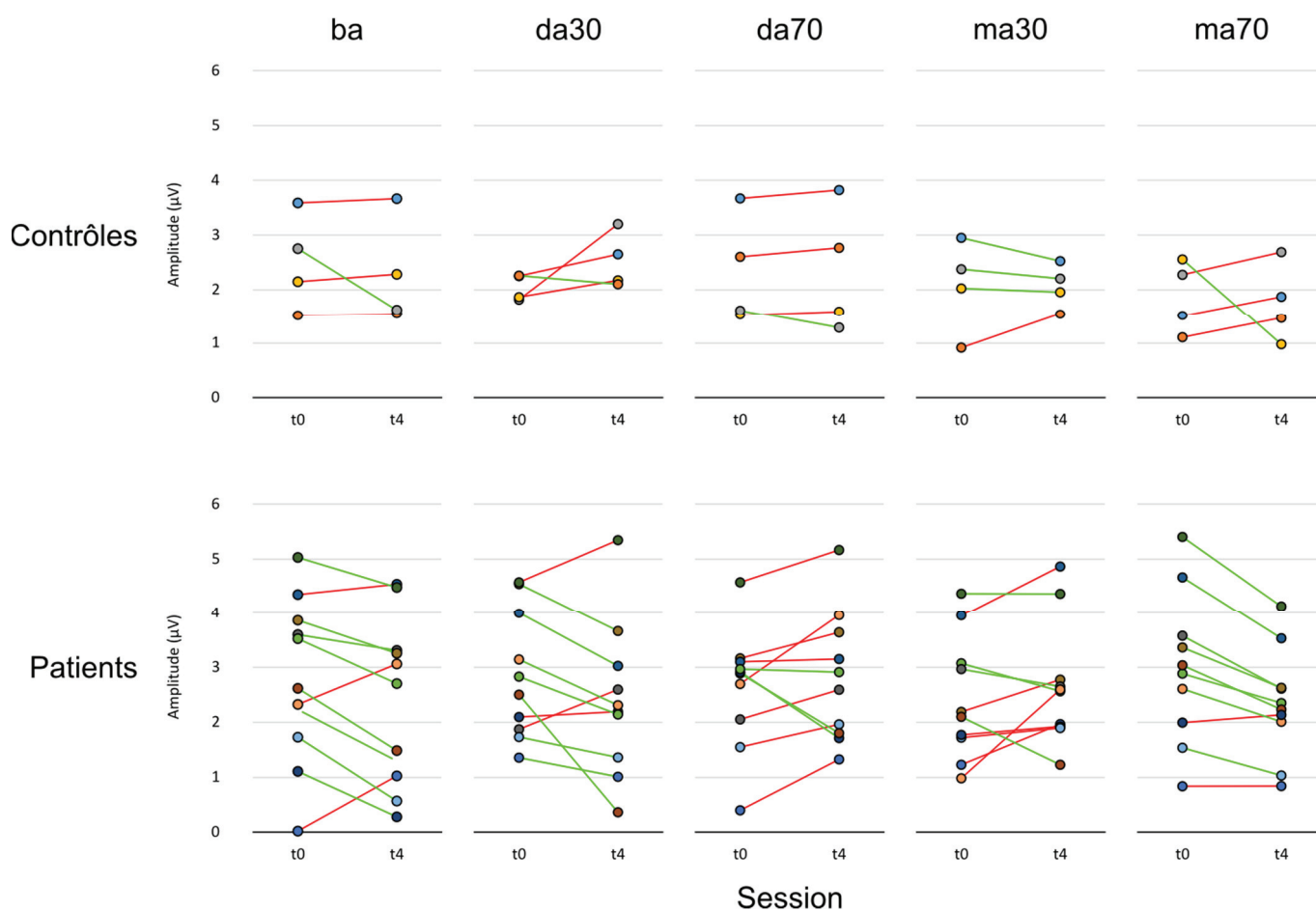


Figure 18. Trajectoires évolutives de la différence d'amplitude N1r-P2r entre t_0 et t_4 , en fonction du groupe et du stimulus. Les couleurs vertes et rouges indiquent respectivement une diminution et une augmentation d'amplitude.

A l'échelle de chaque participant, pour *ma70*, 8 patients sur 10 ont une amplitude différentielle N1r-P2r en baisse entre *t0* et *t4*, contre 1 sur 3 pour les contrôles. Un test de Khi-deux réalisé sur ces proportions donne une significativité au niveau de la tendance ($\text{Khi}^2 = 3.764$; $p = 0.052$), qui suggère ainsi un effet propre aux patients (Figure 18). Les proportions diminution vs. augmentation ont été également testées pour les autres stimuli, sans que cela ne mette en évidence de différence significative ni de tendance ($\text{Khi}^2 = [2.363, 2.363, 0.035, 1.4]$; $p = [0.124, 0.124, 0.852, 0.237]$; respectivement pour *ba*, *da30*, *da70* et *ma30*). Ensemble, ces résultats suggèrent un effet longitudinal propre aux patients et au stimulus *ma70* sur la différence d'amplitude N1r-P2r.

Discussion

Synthèse des résultats.

Le présent chapitre offre une première plongée dans la grande quantité de données acquises durant le protocole, en se concentrant sur l'évolution des mesures pour lesquelles une altération est observée chez les patients malentendants. Ce choix n'exclut bien entendu pas de futures investigations basées sur des variables ne présentant pas de différence entre contrôles et patients.

L'analyse des données d'identification phonémique à l'aide des variables que sont la frontière catégorielle, la pente (ou *zone d'ambiguïté* dans notre cas) et les temps de réponse, a mis au jour plusieurs résultats d'intérêt. Ainsi, les deux groupes *malentendants* et *contrôles* présentent une plus grande difficulté à identifier précisément les syllabes du *continuum* /ba-/ma/ par rapport à celles du /ba-/da/, avec des performances davantage amoindries chez les patients sur le *continuum* /ba-/ma/ et ayant tendance à s'améliorer au fil de l'utilisation des aides auditives. La frontière catégorielle du *continuum* /ba-/ma/ adopte des valeurs différentes entre les groupes ; cependant, elle se déplace vers la droite dans les deux groupes, soit en faveur de la syllabe /ba/, entre les sessions *t0* et *t4*. Enfin, les temps de réponse sont plus lents pour l'identification des syllabes proches de la frontière catégorielle, et une diminution des temps de réponse aux stimuli non-ambigus situés hors de la zone frontière ne s'observe au cours des sessions que pour *continuum* /ba-/da/, chez les patients.

Les réponses électrophysiologiques, à savoir la Speech ABR (*envelope FFR*) et les PEA corticaux, permettent elles aussi plusieurs observations intéressantes. De façon descriptive, la FFR des patients encode davantage la fréquence fondamentale des stimuli par rapport aux contrôles, et les premières harmoniques mettent un temps plus grand à

démarrer leur activité de suivi de phase, entraînant des représentations neuronales dégradées, et par conséquent bien moins distinctes, des différentes syllabes. De même, les réponses transitoires à l'explosion des consonnes (*onset response*) reflètent les difficultés d'identification des différentes syllabes par les patients, ces réponses étant chez eux très similaires pour les 5 stimuli testés, contrairement à ce qui est observé chez les contrôles. Les PEA corticaux présentent également un tel type de perte de distinction entre les syllabes au niveau des paramètres de latence et d'amplitude du complexe N1r-P2r chez les patients. Enfin, au cours des sessions, la surreprésentation des basses fréquences dans la FFR des patients tend à se réguler, et l'amplitude différentielle entre N1r et P2r, anormalement élevée dans ce même groupe, décroît particulièrement pour le *ma70*.

Interprétations croisées.

Continuum /ba/-/ma/ et plasticité sélective.

Des relations peuvent être établies entre ces résultats issus de différents niveaux d'observation. Les patients ont plus de difficultés à identifier /ba/-/ma/ que les contrôles, ces mêmes patients possédant une représentation neuronale moins distinguable entre les différentes syllabes au niveau des Speech ABR comme des PEA corticaux. Bien qu'enregistrés lors d'une tâche d'écoute passive et non durant l'identification phonémique, les PEA corticaux voient l'amplitude N1r-P2r diminuer spécifiquement pour le stimulus *ma70* au fil des sessions, parallèlement à une amélioration de la précision d'identification du *continuum* /ba/-/ma/. La baisse d'amplitude observée dans les réponses corticales pourrait ainsi traduire une identification plus automatique, demandant moins d'effort cognitif, et ainsi sous-tendre l'amélioration comportementale. Une hypothèse alternative et non-exclusive serait l'augmentation de l'amplitude de l'onde P2v, antagoniste de la N1r et la faisant donc diminuer.

Le fait que l'identification ne varie entre les groupes qu'au niveau du *continuum* /ba/-/ma/ suggère que cette faculté liée à la perception catégorielle n'est pas atteinte en soi chez les patients, mais plus vraisemblablement que certains traits phonétiques sont moins bien perçus. Ainsi, la différence de performances d'identification entre les deux *continua* peut être interprétée en se basant sur la qualité des indices acoustiques pertinents pour la distinction de /ba/ et /da/ d'une part, et /ba/ et /ma/ d'autre part. Comme illustré dans la Figure 12, les syllabes /ba/ et /da/ se différencient principalement en un indice spectro-temporel : la transition du deuxième formant, adoptant une trajectoire ascendante pour le /ba/ et descendante pour le /da/. La quasi-unicité de ce caractère distinctif et sa présence à des fréquences relativement bien conservées (en dessous de 2 kHz) peuvent constituer une explication à l'absence de

différence entre patients et contrôles. Entre /ba/ et /ma/, les différences sont plus éparses, avec un contenu spectral plus riche en hautes-fréquences et une intensité plus forte pour le murmure nasal relativement au voisement ; or, la perception des hautes fréquences et la discrimination en intensité sont deux facultés impactées par la surdité neurosensorielle, ce qui peut expliquer une démarcation des performances des patients, dégradées par rapport à celles des contrôles.

Nous pouvons alors interpréter le fait que les effets observés dans la présente étude concernent /ba/-/ma/ davantage que /ba/-/da/ comme une conséquence du constat précédent : la plasticité de réhabilitation s’observerait donc préférentiellement à l’endroit des déficits. Le *continuum* /ba/-/da/ n’est cependant pas dépourvu de tout effet, puisque les temps de réaction s’y améliorent longitudinalement pour les stimuli de la zone hors-frontière. Peut-être est-ce lié à leur meilleure identification, rendant la tâche plus aisée et plus automatique sauf pour quelques syllabes restant ambiguës. Cependant, aucun lien évident n’apparaît avec l’électrophysiologie, les réponses concernées ne présentant pas de spécificité particulière aux syllabes /da/ dans les effets étudiés. Une mesure de l’amplitude de l’*onset response* au *da70* chez les patients pourrait éventuellement être mise en relation avec le comportement, et sera donc implémentée dans la prochaine vague d’analyses.

Référentiel perceptif et variabilité des mesures.

Au niveau de la frontière perceptive des *continua*, un même comportement s’observe dans nos deux groupes, avec un déplacement vers la droite (plus de /ba/ perçus) entre *t0* et *t1* suivi d’une stabilité à travers les sessions suivantes. Les valeurs absolues sont cependant différentes entre les groupes, ce qui est difficile à interpréter dans un effectif aussi restreint : le référentiel perceptif, sorte d’étiquetage automatique des syllabes propre à chaque participant, pourrait être en cause ; mais un effet associé à la presbyacousie pourrait aussi changer la position de la frontière. Une relation serait intéressante à faire entre identité perceptive de chaque stimulus et réponses électrophysiologiques associées : un stimulus identifié à 30% comme /da/ à *t0* sera-t-elle observable pareillement au niveau électrophysiologique que le même stimulus à *t4* identifié alors à 0% comme /da/ suite à un déplacement de la frontière ? Cette question sera également traitée dans les prochaines analyses, avec deux hypothèses possibles : soit le lien perception-encodage neuronal sera dégradé chez les patients, leurs réponses électrophysiologiques étant moins détaillées que celles des contrôles, soit il sera amélioré du fait d’un encodage plus intrinsèque, soit davantage basé sur les voies descendantes, pour compenser le manque d’information en entrée du système nerveux auditif.

En règle générale, nous avons observé une grande variabilité tant entre les réponses des différents participants, comportementales et électrophysiologiques (en considérant l'activité évoquée par le *ba*, physiquement identique chez tous les participants), qu'entre les réponses d'un même participant avec la présence d'un assez fort bruit expérimental lié à la session (état de fatigue, niveau attentionnel, température ambiante, pose des électrodes, entre autres dans la liste des multiples facteurs potentiellement en action). L'utilisation d'un modèle statistique permettant de mieux prendre en compte ce facteur aléatoire permettrait alors peut-être d'approcher des effets plus fins et actuellement masqués par ce bruit.

Fenêtre temporelle longitudinale.

Nous avons ici observé des effets longitudinaux significatifs à la fois sur nos variables comportementales et électrophysiologiques, et ce après seulement 4 mois d'utilisation d'aides auditives. En comparaison à la littérature sur la plasticité de réhabilitation et l'effet d'acclimatation, nous retrouvons l'amélioration des performances cognitives de l'identification phonémique après une même durée d'appareillage, qui avait cependant été mise en évidence sur un *continuum* de voisement, et seulement pour l'oreille droite (Philibert *et al.*, 2003). Nos résultats électrophysiologiques dépassent cependant ce qui avait été observé jusqu'à présent : deux études longitudinales sur une période de 3 mois d'appareillage n'avaient pas pu montrer d'évolution de l'amplitude différentielle des ondes V-SN₁₀ des PEAP, ni de l'amplitude de l'onde P2 des PEA corticaux (Dawes *et al.*, 2013, 2014) ; aussi, une autre étude avait montré un raccourcissement de l'onde V des PEAP, mais au bout de 6 mois d'utilisation d'aides auditives (Philibert *et al.*, 2005) ; une dernière étude avait trouvé une augmentation de l'amplitude de l'onde P2 des PEA corticaux chez des patients appareillés depuis au moins 5 ans (Bertoli *et al.*, 2011). Ainsi, nos effets longitudinaux observés au cours des 4 premiers mois d'appareillage paraissent bien précoces par rapport aux précédentes études, ce qui peut trouver une explication dans notre méthodologie innovante de stimulation à travers les aides auditives fonctionnellement identiques à celles des patients (Bellier *et al.*, 2015b) ; une étude antérieure a en effet montré que des bénéfices acquis dans le référentiel auditif donné par les appareils utilisés au quotidien n'étaient pas révélés par des tests utilisant un autre référentiel auditif pour la stimulation (Gatehouse, 1993). D'autres éléments d'explication sont l'utilisation de stimuli synthétiques aux caractéristiques spectro-temporelles finement contrôlées, la multiplicité des mesures et des sessions d'enregistrement, et assurément la qualité de l'adaptation et de l'affinement des réglages des aides auditives par les audioprothésistes partenaires, condition *sine qua non* d'une remédiation efficace.

Apports potentiels à l'audioprothèse.

Les effets comportementaux mis en évidence dans cet article, et assimilables aux résultats d'une étude antérieure montrant une amélioration de l'identification phonémique au bout de 4 mois d'appareillage (Philibert *et al.*, 2003), suggèrent un apport potentiel de l'identification phonémique dans la caractérisation des déficits occasionnés par les troubles auditifs. Effectivement, l'audiométrie vocale permet à l'heure actuelle de quantifier les confusions phonémiques pour donner un indice d'intelligibilité de la parole : un taux d'erreur bas sera interprété comme une compréhension aisée. Or, certaines listes utilisées utilisent des mots dotés de sens (contrairement aux logatomes d'autres listes), si bien que les voies corticofuges efférentes (*top-down*) peuvent compenser une carence dans la représentation neuronale bas-niveau des phonèmes perçus, le test étant alors moins spécifique de la perception auditive pure. Aussi, les mots sont dans bien des cas produits par un même locuteur, avec en conséquence un jeu de paramètres spectro-temporels relativement restreint pour un même phonème ; cependant, les patients se plaignent souvent de moins bien entendre certaines voix (de femmes ou d'enfants le plus souvent). De fait, l'audiométrie vocale teste la reconnaissance binaire d'un phonème donné, là où le test d'identification phonémique permet d'estimer la plage des paramètres spectro-temporels rendant possible la perception de ce même phonème. Un inconvénient est néanmoins la durée d'un tel test (environ 5 minutes par phonème), difficile à réaliser sur un grand ensemble de sons de parole ; cette limite pourrait alors être contournée par un ciblage de phonèmes particulièrement sensibles, qui devront être définis par des études futures.

Concernant l'électrophysiologie, la Speech ABR a été promise à un avenir prometteur dans la prise en charge des troubles de l'audition, qu'il s'agisse : de la caractérisation de ces troubles pour la sélection d'aides auditives appropriées ; de leurs réglages basés sur une mesure objective plus proche de l'intelligibilité de la parole que ne l'est l'audiométrie tonale ; de l'estimation finement quantifiée des bénéfices perceptifs ; ou encore de la prédiction de succès d'un appareillage auditif par une identification des capacités d'encodage spectro-temporel de chaque patient (Dajani *et al.*, 2013 ; Anderson & Kraus, 2013 ; Bellier *et al.*, 2014). La présente étude, apportant les premières Speech ABR obtenues *via* des aides auditives chez le patient malentendant, constitue de fait un premier pas vers l'audiométrie objective basée sur l'encodage neuronal des sons de parole. La temporalité des différents effets des aides auditives sur la Speech ABR sont toutefois à prendre en compte : une mesure comparative entre la séance d'adaptation et quelques mois plus tard permettra éventuellement une quantification des bénéfices de l'appareillage, mais ne présentera pas un intérêt opérationnel fort pour le réglage des aides auditives, puisque les réglages prothétiques ont besoin d'être affinés à plusieurs reprises au cours des premiers mois (tout du moins dans notre conception actuelle). Par

contre, l'établissement de mesures normatives permettrait d'affiner les réglages à chaque session jusqu'à ce qu'une composante donnée de la Speech ABR atteigne une zone cible de valeurs, avec pour limitation le temps d'acquisition de cette mesure d'environ 10 minutes pour une seule syllabe. Quoiqu'il en soit, les techniques d'analyse temps-réel du signal EEG sont en rapide évolution, si bien que le temps requis pour mesurer une Speech ABR diminuera très probablement dans les années à venir.

Prochaines analyses.

Comme nous avons pris le soin de le détailler dans la partie Matériel et méthodes, le protocole de cette étude a permis de recueillir une grande quantité de données de natures différentes, pour lesquelles seule une partie des analyses possibles a été présentée ici. De futurs traitements et analyses supplémentaires seront donc requis afin d'exploiter toute la richesse de ces données.

Sur la préparation des données avant extraction des mesures, nous pouvons notamment affiner nos techniques de traitement des données électrophysiologiques pour débruiter les Speech ABR de participants présentant un artéfact musculaire post-auriculaire, comme c'est le cas de P10 pour toutes ses sessions, et ainsi gagner en puissance statistique. Aussi, un artéfact de stimulation a été observé sporadiquement, chez quelques patients et à des sessions isolées, dans la Speech ABR en polarités séparées ; la correction de cet artéfact permettra d'étudier pleinement la FFR spectrale, issue de la soustraction des réponses aux stimuli de polarités opposées (et de la division par deux du signal résultant), et focalisée sur la partie haute du spectre. Il serait alors possible d'explorer l'encodage en suivi de phase des deux premiers formants dans les voies auditives nerveuses (Plyler & Ananthanaryan, 2001), avec pour limite la fréquence anti-repliement fixée à 1000 Hz de notre système d'acquisition.

De nombreuses mesures peuvent être extraites des différentes réponses comportementales et électrophysiologiques ; ces potentialités sont détaillées ci-dessous.

- Du point de vue de l'audiologie, les analyses prévues s'intéresseront à la latence et l'amplitude des PEAP de $t-1$ pour comparaison aux autres variables dépendantes ; à la comparaison de l'intensité de stimulation apportant la même sonie entre $t0$ et $t4$, pour mettre en lumière un possible effet longitudinal ; à l'exploitation des questionnaires remplis à $t-1$ avec notamment la durée de privation auditive, prédicteur potentiel de la taille des effets longitudinaux dans le sens où les voies auditives d'un patient atteint d'une surdité ancienne et bien ancrée seront moins aisément malléables par rapport à une surdité d'installation plus récente ; à la quantification des pas de réglage prothétique,

pour corrélérer le degré de changement acoustique apporté par les aides auditives aux effets longitudinaux observés sur nos différentes variables.

- Pour l'identification phonémique, nous avons déjà parlé de la possibilité de mesurer les temps de réponse à la frontière en indexant la taille de la zone frontière sur la taille de la zone d'ambiguïté liée à la pente de la courbe sigmoïde. Nous pourrions également corrélérer les différentes variables, pour comprendre par exemple les influences croisées entre pente d'identification et mouvement de la frontière catégorielle. Les confusions phonémiques mesurées par l'audiométrie vocale pourraient aussi être comparées aux résultats pour les *continua* /ba/-/da/ et /ba/-/ma/, afin de tenter l'établissement d'un lien entre les deux tests. Enfin, nous pourrions prendre en compte l'identité de chaque syllabe dans les *continua*, pour tenter de comprendre les différences interindividuelles des données électrophysiologiques : un patient avec stimulus *ma70* défini à la syllabe 30 du *continuum* /ba/-/ma/ n'aura certainement pas les mêmes activités évoquées qu'un autre patient ayant pour *ma70* la syllabe 48, par exemple.
- Les Speech ABR offrent une mesure fine de l'activité neuronale en réponse aux sons de parole, et pourraient alors servir à caractériser un déficit de synchronisation neuronale, par l'extraction d'une mesure de consistance des essais (*phase-locking value*). Les différences entre fronts de montée des patients et contrôles dans les profils spectraux (Figure 11) est quantifiable par calcul de l'aire entre les deux courbes, pour que nous soyons à mêmes de cibler cette valeur dans son évolution longitudinale potentielle. Aussi, des tests statistiques tels que le Kruskal-Wallis peuvent révéler des différences entre Speech ABR des patients et contrôles, comme par exemple au niveau de l'*onset response*.
- Concernant les PEA corticaux, nous nous sommes ici concentrés sur les latences, amplitudes propres et amplitudes différentielles des ondes en présence ; nous pourrions y ajouter la mesure de leurs latences différentielles pour compléter le relevé. Aussi, la comparaison des PEA corticaux (mais aussi des résultats de l'identification phonémique) en fonction des réglages prothétiques (t_0 ou t_4 ; les deux étant testés à t_4) peut apporter des informations d'intérêt sur les liens entre acoustique, encodage neuronal et perception. Aussi valable pour les Speech ABR, l'étude des topographies peut enfin apporter une information quant aux générateurs sujets à la plasticité neuronale, telle qu'observable par exemple dans la baisse de l'amplitude différentielle N1r-P2r.
- Enfin, l'ensemble de ces mesures pourront être combinées dans un modèle statistique plus sophistiqué que les tests effectués ici, afin de prendre en compte le plus grand nombre de sources de variance possible, avec pour limite la taille réduite de notre effectif.

Conclusion.

Le présent chapitre présente les premiers résultats de l'étude menée chez le malentendant appareillé. En combinant mesures comportementales et électrophysiologiques sur quatre mois d'enregistrements, dont la Speech ABR jusqu'alors jamais rapportée chez le patient malentendant appareillé, il apporte une lumière nouvelle sur la plasticité de réhabilitation consécutive à l'utilisation d'aides auditives. Ainsi, une amélioration de l'identification du trait de nasalité est mise en évidence, en parallèle d'une baisse en amplitude de deux ondes des PEA corticaux en réponse aux syllabes nasales. La Speech ABR voit quant à elle l'amplitude de sa fréquence fondamentale ainsi que son SNR diminuer au fil des sessions, allant donc dans le sens d'un rééquilibrage de l'encodage basses vs. hautes-fréquences et donc d'une amélioration dans l'intelligibilité de la parole.

Matériel et méthodes

Participants. Dix patients malentendants presbycousiques (61 à 80 ans, moyenne de 71 avec écart-type de 7 ; 3 femmes) et quatre individus normo-entendants (62 à 65 ans, moyenne de 64 avec écart-type de 1 ; 2 femmes) ont participé à la présente étude. Les patients nous ont été adressés par le laboratoire d'audioprothèse *Voir et Entendre* (Lyon, France). Pour être inclus dans le groupe *malentendants*, un patient devait : présenter une surdité bilatérale et symétrique (à plus ou moins 20 dB) de type presbycousie, soit avec basses fréquences (inférieures à 1 kHz) bien conservées et pertes dans les hautes fréquences ; ne jamais avoir été appareillé auparavant, pour potentialiser la plasticité de réhabilitation ; et être appareillé avec des aides auditives *Phonak Audéo Q90-312*, que nous avons auparavant minutieusement contrôlé dans leur capacité à évoquer la Speech ABR. Les candidats pour l'inclusion au groupe contrôle *normo-entendants* devaient ne présenter aucune perte au-delà de 30 dB HL sur les fréquences de 125 Hz à 3 kHz. Cette marge de tolérance (le seuil normo-entendant étant usuellement fixé à 20 dB HL pour les fréquences entre 250 Hz et 8 kHz) a été définie pour pallier à la rareté des profils audiométriques strictement normo-entendants passé la soixantaine (Gates & Mills, 2005). L'existence de troubles du langage de type dyslexie, neurologiques ou psychiatriques constituait un critère de non-inclusion des deux groupes. Chaque participant devait enfin donner son consentement éclairé avant de commencer le protocole, dont le contenu avait été préalablement approuvé par le Comité de Protection des Personnes Lyon SUD-EST IV.

Protocole longitudinal. La présente étude a été conçue le long d'un axe temporel long de 4 mois, avec 5 sessions d'enregistrement pour chaque participant, ainsi qu'une session préliminaire. Celle-ci, appelée *t-1*, était dédiée à l'explication détaillée du protocole, à la signature du consentement éclairé, et le cas échéant au remplissage de questionnaires et à l'enregistrement des PEAP. La première des 5 sessions d'enregistrement, appelée *t0*, correspondait au jour de l'adaptation des aides auditives par l'audioprothésiste pour les patients presbycusiques, et était fixée arbitrairement pour les participants du groupe contrôle. Les quatre sessions suivantes (*t1*, *t2*, *t3* et *t4*) se déroulaient respectivement deux semaines, un, deux et quatre mois après *t0*. Ces cinq sessions d'enregistrement partageaient toutes un même tronc commun, composé d'une tâche comportementale d'identification phonémique de syllabes s'opposant en lieu d'articulation et en nasalité, et un enregistrement de la Speech ABR et des PEA corticaux auxdites syllabes ; seules *t0* et *t4* variaient de ce tronc commun par la présence de tâches supplémentaires. La taille de notre fenêtre temporelle d'observation a été principalement choisie sur la base de précédentes études montrant l'existence d'une plasticité de réhabilitation après les premiers mois d'utilisation d'aides auditives (Philibert *et al.*, 2003, 2005 ; Munro, 2008), ainsi qu'en prenant en compte l'expérience d'audioprothésistes chevronnés vis-à-vis de la cinétique d'amélioration de la perception de parole, telle que mesurée par l'audiométrie vocale.

Syllabes. *Rationnel.* Les traits phonétiques de lieu d'articulation et de nasalité sont particulièrement sujets à confusion pour les malentendants (Lefèvre, 2006). Aussi, la probabilité d'observer un effet d'acclimatation, agissant sur leur perception ou leur encodage neuronal suite à l'amplification des indices pertinents par les aides auditives, est donc plus grande. Dans cette optique, et pour être capables d'évaluer les performances d'identification phonémique des participants, nous avons conceptualisé un carré de stimuli avec les syllabes /ba/ (bilabiale orale), /da/ (dentale-alvéolaire orale), /ma/ (bilabiale nasale) et /na/ (dentale-alvéolaire nasal) aux quatre angles. Les côtés étaient alors matérialisés par des *continua* acoustiques, qui sont aux sons ce que les dégradés sont aux couleurs. Du fait des contraintes de durée des sessions expérimentales et de nos expériences préalables avec la syllabe /ba/ pour évoquer la Speech ABR (Bellier *et al.*, 2013, 2015a, 2015b), nous avons décidé de n'étudier que les côtés /ba/-/da/ et /ba/-/ma/ de notre carré de syllabes (s'opposant respectivement en lieu d'articulation et nasalité).

Synthèse. La qualité des réponses électrophysiologiques est fondamentalement déterminée par la structure spectro-temporelle des stimuli permettant de les évoquer, ce qui est d'autant plus vrai pour la Speech ABR (Skoe & Kraus, 2010). De ce fait, nous avons opté pour l'utilisation de parole synthétique. Les syllabes /ba/, /da/ et /ma/ ont ainsi été générées en utilisant le synthétiseur formantique de René Carré, qui offre un contrôle relativement fin des paramètres spectro-temporels des sons de parole, obtenus ici sur la base de l'analyse acoustique de syllabes naturelles enregistrées auprès de différents locuteurs. Des *continua* /ba/-/da/ et /ba/-/ma/ de 50 stimuli chacun ont ensuite été obtenus par interpolation à partir des syllabes « pures ».

Caractéristiques spectro-temporelles. Les syllabes pures /ba/, /da/ et /ma/ étant toutes trois des occlusives voisées, elles partagent la même structure *voisement-relâchement-voyelle*. Leur durée totale a été fixée à 180 ms, et leur fréquence fondamentale (F0) à 170 Hz, correspondant à une voix grave de femme. Le voisement des consonnes /b/ et /d/ (ou le murmure nasal du /m/) s'étale sur les 60 premières millisecondes du signal ; il est suivi par une transition formantique de 30 ms correspondant au relâchement de l'occlusion (*release*), menant à la voyelle /a/ qui occupe les 90 dernières millisecondes. Si les deux premières phases sont variables, la dernière est strictement identique entre les trois syllabes (et donc entre tous les éléments des *continua* /ba/-/da/ et /ba/-/ma/). Concernant le /ba/, les formants du voisement oscillent autour de 160, 950, 2900 et 4250 Hz (pour F1, F2, F3 et F4), puis F1 et F2 varient au cours des 30 ms de transition jusqu'à atteindre leur cible, à 825 et 1360 Hz respectivement ; ces dernières valeurs sont enfin maintenues pendant les 90 ms du /a/. La syllabe /da/ ne varie du /ba/ que sur F2, qui assure une transition formantique descendante de 1870 à 1360 Hz entre voisement et voyelle (Liberman *et al.*, 1954). Enfin, le murmure nasal du /ma/ est plus ample que le voisement des autres consonnes, et ses formants varient autour de 215, 1400, 2550 et 4250 Hz ; le vide énergétique ainsi approximé à 1 et 3 kHz mime les anti-formants, indices de nasalité dus au manque de résonance de la cavité orale (Johnson, 2003).

Sélection des syllabes pour l'EEG. Avec les paramètres de stimulation que nous avons définis (voir ci-dessous), la Speech ABR évoquée par une syllabe s'enregistre en environ 10 minutes, et les PEA corticaux en environ 2 minutes. Pour ne pas dépasser une heure d'enregistrement, déjà précédée de la partie comportementale et de l'installation des électrodes, nous avons décidé de limiter nos mesures à 5 stimuli. Leur sélection se basait sur les résultats des tâches d'identification phonémique sur les *continua* /ba/-/da/ et /ba/-/ma/ à *t0*, et dépendait donc de chaque participant. Le /ba/ pur, partagé par les deux *continua* et ayant le plus de chances d'être identifié comme /ba/ de façon constante à travers les sessions, était toujours sélectionné. Puis dans chacun des *continua*, les deux stimuli perçus à 30 et 70% comme /da/ ou /ma/, et dont la perception est susceptible

de changer au fur et à mesure de l'appareillage, étaient identifiés. Les 5 syllabes ainsi déterminées, propres à chaque participant, et par la suite désignées comme *ba*, *da30*, *da70*, *ma30* et *ma70*, étaient ensuite utilisées pour la partie électrophysiologie de *t0* et des quatre sessions suivantes.

Stimulation. Les stimuli sont délivrés de façon binaurale *via* des appareils auditifs *Phonak Audéo Q90-312* techniquement modifiés pour notre protocole : au lieu d'être capté par les micros, le signal acoustique entre directement dans l'appareil par voie filaire, ce qui permet d'assurer une latence hautement reproductible, indispensable à l'obtention de la Speech ABR. En effet, les autres possibilités auraient été de jouer le signal en champ libre pour qu'il soit naturellement capté par les micros des appareils, mais nous aurions risqué un déphasage des essais unitaires de la Speech ABR à cause des mouvements de tête et changements de position des participants au cours du temps d'enregistrement ; ou de transmettre le signal aux appareils en mode sans-fil (type Bluetooth), mais les tests préliminaires effectués ont révélé que nous aurions perdu beaucoup en fiabilité temporelle. Au niveau des écouteurs déportés des *Audéo Q90-312* (situés dans le conduit auditif externe plutôt que sur l'appareil même), nous avons adapté des embouts en mousse (de type inserts) pour que les basses fréquences délivrées ne fuient pas du conduit auditif (Bellier *et al.*, 2015b). Concernant les patients malentendants, les réglages personnalisés (gains et facteurs de compression, principalement) définis par l'audioprothésiste lors de l'adaptation des aides auditives à *t0* étaient copiés à chaque session dans nos aides auditives modifiées ; les contrôles normo-entendants ont quant à eux tous eu les mêmes réglages prothétiques, calculés pour un audiogramme tonal plat à 30 dB HL. De plus, le volume sonore était fixé à 60 dB SL grâce à une tâche comportementale itérative réalisée à *t0* et permettant de déterminer l'intensité de stimulation nécessaire à l'atteinte du seuil de perception et donc propre à chaque participant. Cette valeur d'intensité était ensuite utilisée pour les mesures comportementales et électrophysiologiques de *t0* et des sessions suivantes. Ces deux points, à savoir l'utilisation d'une même intensité de stimulation et des mêmes réglages prothétiques à travers toutes les sessions, étaient absolument indispensables pour assurer une constance dans la sortie acoustique des aides auditives : une éventuelle évolution dans les mesures comportementales ou électrophysiologiques peut alors s'interpréter comme une conséquence de la plasticité de réhabilitation et non comme un simple effet du changement de signal acoustique en entrée du système auditif.

Particularités à *t0* et *t4*. Toutes les sessions à partir de *t0* partagent le même noyau : tâche comportementale d'identification phonémique pour les *continua* /ba/-/da/ et /ba/-

/ma/, et recueil électrophysiologique de la Speech ABR et des PEA corticaux. Cependant, la session $t0$ était précédée par la simplification des réglages prothétiques de chaque malentendant avant leur copie dans les aides auditives modifiées (désactivation des algorithmes de traitement du signal et des programmes superflus), et par la tâche de calibration du volume sonore décrite précédemment. A $t4$, cette même tâche était réitérée dans le simple but de pouvoir comparer l’audibilité par rapport à $t0$, la valeur d’intensité obtenue n’étant pas utilisée par la suite. Aussi, la tâche d’identification phonémique pour les deux *continua* et la mesure des PEAc étaient reproduites avec les réglages prothétiques les plus récents, et ce bien entendu uniquement pour le groupe *malentendants* (puisque les réglages *normo-entendants*, calculés pour une courbe audiométrique plate à 30 dB et partagés par tous les participants du groupe *contrôle*, n’évoluaient pas au cours des sessions). Ces mesures supplémentaires apportent une estimation de la plus-value des derniers réglages prothétiques, à moindre coût en termes de temps expérimental ajouté.

Questionnaires. Lors de la session de recrutement $t-1$, les participants ayant donné leur consentement éclairé remplissaient deux questionnaires. Le premier permettait de récolter des informations générales (âge, genre, niveau d’études, situation professionnelle, langue maternelle), concernant le niveau d’expertise auditive (pratiques linguistiques et musicales) ou le passé audiolgique (environnement de travail bruyant, traumatismes sonores, acouphènes, orthophonie, troubles du langage). Le second questionnaire était le bien connu *Edinburgh Handedness Inventory* (Oldfield, 1971), permettant de quantifier la latéralité manuelle.

PEAP. Consécutivement au renseignement des questionnaires de $t-1$, les seuils des PEAP étaient mesurés pour chacune des deux oreilles, à l’aide du *Centor USB* (fréquence d’échantillonnage de 13,3 kHz ; *Deltamed*, France) et de ses 4 électrodes placées au sommet du front (Fpz), sur les lobes d’oreille (référence, ipsilatérale à l’oreille stimulée) et sur la pommette droite (masse). Des clics de raréfaction (polarité unique) d’une durée de 100 μ s étaient délivrés à une cadence de 19 clics par seconde, et 1600 essais étaient enregistrés avec un filtre passe-haut doté d’une fréquence de coupure de 70 Hz. Afin de renforcer la fiabilité des mesures, deux tracés par oreille et intensité de stimulation étaient enregistrés en alternant oreille droite et oreille gauche, et en partant de la plus haute intensité supportable pour descendre par pas de 10 dB nHL jusqu’à disparition de l’onde V. L’enregistrement de ces tracés de PEAP durait entre 20 et 40 minutes, durant

lesquelles le participant était confortablement installé dans un fauteuil en regardant un film sous-titré et sans bande son.

Calibration de l'intensité. A t_0 , une tâche de calibration du volume de stimulation était effectuée. La syllabe *ba* était délivrée en association à un indice visuel, suite à quoi le participant devait indiquer s'il l'avait perçue (sans l'avoir nécessairement comprise). Le cas échéant, le volume sonore était réduit, et dans le cas contraire, augmenté, avant une nouvelle présentation de la syllabe ; cette séquence se répétait ensuite jusqu'à atteinte du seuil de perception. Pour toutes les sessions, les stimuli utilisés dans les tâches comportementales et les enregistrements électrophysiologiques étaient délivrés à 60 dB SL, par rapport au seuil mesuré à t_0 . A t_4 , la tâche de calibration avec les réglages prothétiques de t_0 était reproposée, pour prendre mesure d'une possible évolution dans l'audibilité des *stimuli* ; aussi, et seulement pour le groupe *malentendants*, cette calibration était effectuée pour les aides auditives programmées avec les réglages les plus récents, afin de déterminer le volume sonore à utiliser pour les tests supplémentaires d'identification phonémique et de PEA corticaux réalisés avec ces nouveaux réglages (voir ci-dessus).

Identification phonémique. *Déroulement.* La tâche comportementale d'identification phonémique était proposée pour les deux *continua* /ba/-/da/ et /ba/-/ma/. Elle comportait un court entraînement, suivi du recueil des réponses d'identification. L'entraînement était répété à chaque session et avant chaque *continuum*, dans le but de familiariser le participant aux commandes de la souris et à l'acoustique des *stimuli*. L'ordre des *continua* et l'assignement des deux boutons de la souris aux réponses étaient déterminés aléatoirement et contrebalancés entre les participants, mais restaient constants au fur et à mesure des sessions pour un même participant. Les *stimuli* étaient délivrés à l'intensité calibrée à t_0 via les aides auditives modifiées programmées avec les réglages de chaque participant. Le déroulement de cette tâche était contrôlé par le logiciel *Presentation* (Neurobehavioral Systems, USA). La suite de la description concerne de la même manière les deux *continua*.

Entraînement. Onze syllabes allant d'un extrême à l'autre (par exemple de *ba* à *da*) étaient délivrées au participant puis, après une courte pause, était présentées dans l'ordre inverse (de *da* vers *ba*). Après chaque présentation, le participant devait indiquer ce qu'il avait perçu en cliquant sur l'un ou l'autre des boutons de la souris (par exemple clic gauche pour /ba/ et droit pour /da/). Cette phase d'entraînement pouvait être répétée jusqu'à ce que le participant se sente à l'aise avec la tâche et les syllabes des *continua*.

Tâche. Les 50 *stimuli* n'étaient présentés qu'une seule fois chacun, au travers de 5 blocs de 10 séparés par des pauses. Chaque bloc contenait 2 *stimuli* de chaque dizaine (par exemple les syllabes 2 et 7, 15 et 18, 21 et 29...), de manière à ce qu'il ne puisse y avoir que des syllabes perçues comme /ba/, par exemple, et l'ordre était randomisé à l'intérieur de chaque bloc. Les participants disposaient de temps de pause réguliers. L'accomplissement des deux séquences *entraînement-tâche* durait environ 15 minutes.

Sélection des syllabes à t0. Les données d'identification phonémique permettent d'obtenir une courbe de perception catégorielle, qui représente le pourcentage d'identification /da/ ou /ma/ en fonction de l'index du stimulus (de 1 à 50). Pour cela, et comme un clic précipité et erroné durant la réalisation de la tâche influencerait grandement les paramètres de cette courbe sigmoïde, nous avons procédé en premier lieu à un rejet des réponses aberrantes (*outliers*), par une procédure itérative où les réponses les plus éloignées de leur valeur théorique selon une distribution binomiale (adoptant les paramètres d'un *modèle linéaire généralisé* adapté aux données) étaient écartées une à une des données, jusqu'à ce que la somme des résidus ne soit plus réduite significativement (comme indiqué par un test de Chi2 pour 1 degré de liberté et un seuil fixé à 0.01 ; Pregibon, 1981 ; Bedrick & Hill, 1990). En second lieu, la courbe de perception catégorielle a été obtenue en adaptant le même précité *modèle linéaire généralisé* (avec distribution d'erreur et fonction de lien binomiales), cette fois sur les données débarrassées des valeurs aberrantes. En projetant sur l'axe des abscisses les points de la sigmoïde à 30 et 70% de reconnaissance /da/ ou /ma/, nous obtenions l'indice des *stimuli* à utiliser durant toute la suite du protocole, pour le participant donné.

Analyses. Les paramètres de la sigmoïde obtenue comme décrit précédemment nous renseignent directement sur les valeurs de pente et de frontière (indice du *stimulus* perçu à 50% comme /da/ ou /ma/). Notons que la pente est directement reliée à la notion de *précision catégorielle* : plus elle est redressée, plus les catégories phonémiques sont distinctes (Simon & Fourcin, 1978) ; la notion de perception catégorielle n'est quant à elle accessible que par la combinaison des résultats aux tâches d'identification et de discrimination phonémique (Liberman *et al.*, 1957), ce qui n'était pas réalisable ici sans augmenter drastiquement la durée des sessions. Les temps de réaction ont également été mesurés pour chaque réponse, et moyennés en deux groupes selon la position du stimulus associés relativement à la frontière catégorielle ; ainsi, le groupe *frontière* associe les 9 stimuli centrés sur la frontière, tandis que le groupe *hors-frontière* représente la moyenne les valeurs des temps de réponse aux 41 stimuli restants.

Electrophysiologie. Après avoir complété la tâche d'identification phonémique pour les deux *continua*, les participants s'asseyaient dans un fauteuil confortable. Les 32 électrodes actives *actiCAP* du système *BrainAmp* (fréquence d'échantillonnage de 5 kHz ; *Brain Products*, Allemagne) leur étaient alors installées, *via* un bonnet adapté à leur périmètre crânien excepté pour TP9 et TP10, directement collées aux mastoïdes gauche et droite (respectivement). La référence était placée au nez, la masse en AFz, et les impédances étaient descendues sous 10 k Ω . Après cette installation d'une durée proche de 30 minutes, les Speech ABR évoquées par les cinq syllabes sélectionnées à *t0* étaient recueillies, suivies par l'enregistrement des PEA corticaux en réponse aux mêmes syllabes. Pour maintenir les participants éveillés durant le recueil de ces deux réponses électrophysiologiques, plusieurs films sous-titrés et diffusés sans le son étaient soumis à leur choix. Enfin, de courtes pauses ponctuaient régulièrement l'enregistrement EEG d'une durée d'environ 1 heure, durant lesquelles des boissons leur étaient proposées.

Speech ABR. Stimulation. Les syllabes *ba*, *da30*, *da70*, *ma30* et *ma70* étaient délivrées dans un ordre aléatoire et de façon binaurale *via* les appareils auditifs modifiés, à l'intensité correspondant à 60 dB SL telle que déterminée à *t0*. La cadence de stimulation était de 3,6 *stimuli* par seconde, avec une variabilité temporelle de 50 ms (*jitter*). Aussi, les syllabes étaient présentées en polarité alternée, pour minimiser l'impact d'un potentiel artefact de stimulation (Campbell *et al.*, 2012). Mille répétitions de chaque polarité pour chaque syllabe étaient enregistrées, pour un total de 10000 essais. L'enregistrement des Speech ABR durait environ 50 minutes.

Traitement des données. Les données brutes étaient d'abord débarrassées des artefacts du secteur électrique (50 Hz et harmoniques impaires) avec une série de filtres *notch*, puis étaient filtrées entre 80 et 1000 Hz (filtre passe-bande Butterworth d'ordre 5) pour isoler l'activité haute-fréquence de la Speech ABR de celle basse-fréquence des PEA corticaux. Les mille essais de chaque polarité pour chaque syllabe étaient ensuite découpés entre -50 et 300 ms par rapport au début des stimuli, puis moyennés avec correction de ligne de base entre -50 et 0 ms et implémentation d'un rejet d'artéfact sur gabarit (le seuil était défini à 2 écarts-types au-delà de la moyenne). Les potentiels évoqués obtenus étaient re-référencés à la moyenne des mastoïdes, et les réponses aux polarités opposées étaient enfin moyennées. Cette dernière étape permet d'obtenir la FFR d'enveloppe, tandis que la FFR spectrale s'obtiendrait en moyennant la réponse à une polarité avec l'opposé de la réponse à l'autre polarité ; la première privilégiant les basses fréquences (F0 et premières harmoniques) en minimisant le potentiel artéfact de stimulation, et la seconde soulignant les hautes fréquences (au-delà de 500 Hz ; Aiken & Picton, 2008). Dans la première vague d'analyses présentée ici, nous n'avons étudié que

la FFR d'enveloppe, que nous appelons alors Speech ABR par souci de simplification. Les différentes étapes du traitement des données ont été réalisées en utilisant les logiciels Elan (Aguera *et al.*, 2011) et MATLAB (The MathWorks, USA).

Analyses. Dans le cadre de cette thèse, nous présentons une première série d'analyses des données, étudiant la composition spectrale de la Speech ABR pour tester l'hypothèse d'un recul de la surreprésentation des basses fréquences chez le malentendant (Anderson *et al.*, 2013), au fur et à mesure de l'utilisation des aides auditives. Nous avons d'abord calculé le *rapport signal sur bruit* (*Signal-to-Noise Ratio*, SNR) des Speech ABRs pour chaque stimulus et chaque session de chaque participant, cette mesure étant dominée par la réponse aux basses fréquences, ici particulièrement au 170 Hz de la F0. Pour le calcul du SNR, la fenêtre temporelle de *bruit* était définie entre -50 et 0 ms, et celle de *signal + bruit* entre 80 et 130 ms, couvrant ainsi le début de la Frequency-Following Response (FFR). Pour raffiner cette analyses, nous avons également calculé la transformée de Fourier (FFT) des réponses entre 80 et 160 ms, soit durant la FFR évoquée par la fin de la transition formantique et une grande partie de la voyelle /a/. Dans les spectres obtenus, la moyenne quadratique (*RMS*) de l'énergie à une fréquence donnée était d'abord calculée sur une fenêtre d'environ 20 Hz centrée sur cette fréquence, avant d'être normalisée par la moyenne quadratique de l'énergie environnante (fenêtres de -90 à -40 Hz et de 40 à 90 Hz par rapport à la fréquence d'intérêt). Cette mesurée était effectuée pour la fréquence fondamentale et ses 3 premières harmoniques.

PEA corticaux. *Stimulation.* Les PEA corticaux ont été enregistrés avec les mêmes paramètres de stimulation que pour les Speech ABR, à ceci près de la cadence de stimulation (1,1 syllabe par seconde) et du nombre d'essais recueillis (100 par polarité par syllabe, la réponse étant environ 10 fois plus ample). Par conséquent, l'acquisition totale des 1000 essais durait environ 10 minutes.

Traitement des données. Comme pour les Speech ABR, la première étape de traitement était le filtrage des artefacts du secteur. Les données étaient ensuite filtrées entre 3 et 20 Hz (passe-bande Butterworth d'ordre 3), avant que les essais découpés entre -100 et 400 ms ne soient moyennés avec correction de ligne de base de -100 à 0 ms et rejet sur critère de dynamique d'amplitude (seuil à 60 μ V par essai de 500 ms). Enfin, les réponses aux polarités opposées étaient moyennées.

Analyses. Les amplitudes et latences des ondes N1 et P2 en réponse au relâchement de l'occlusion (*release*) ont été mesurées par marquage manuel pour chacune des réponses aux 5 stimuli, à travers les 5 sessions de *t0* à *t4*. Pour ce faire, et comme il existait une grande variabilité entre les PEA corticaux des différents participants sans doute en

grande partie due à une sélection de syllabes *da30*, *da70*, *ma30* et *ma70* différentes pour chacun, les grandes moyennes des réponses évoquées ont été calculées à travers les 5 sessions, et ce pour chaque *stimulus* chez chaque participant. Les latences des ondes moyennes alors observées ont été marquées manuellement pour déterminer les fenêtres temporelles de marquage automatique des latences et amplitudes sur tous les stimuli et toutes les sessions.

Remerciements

Nous tenons à rendre hommage à l'équipe de *Voir et Entendre* (Lyon), laboratoire de l'audition mené par Jean-François Vesson, pour nous avoir adressé les patients malentendants dans d'aussi bonnes conditions. Nous remercions également Willy Serniclaes (Laboratoire de Psychologie de la Perception, CNRS UMR 8158, Paris) pour ses précieux conseils dans la synthèse des stimuli, et Patrick Bouchet pour toute son aide dans la mise en place du protocole.

Références

- Aguera P-E, Jerbi K, Caclin A, Bertrand O (2011) ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Comput Intell Neurosci* 2011:158970.
- Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245(1-2):35–47.
- Albers K (2012) Hearing loss and dementia: new insights. *Minn Med* 95(1):52–54.
- Anderson S, Kraus N (2013) The Potential Role of the cABR in Assessment and Management of Hearing Impairment. *Int J Otolaryngol* 2013:604729.
- Anderson S, Parbery-Clark A, White-Schwoch T, Drehabl S, Kraus N (2013) Effects of hearing loss on the subcortical representation of speech cues. *J Acoust Soc Am* 133(5):3030–3038.
- Arlinger S, et al. (1996) Report of the Eriksholm Workshop on auditory deprivation and acclimatization. *Ear Hear* 17(3 Suppl):87S–98S.
- Bakhos D, Villeuneuve A, Kim S, Hammoudi K, Hommet C (2015) Hearing loss and Alzheimer's disease. *Geriatr Psychol Neuropsychiatr Vieil* 13(2):195–204.
- Bao J, Ohlemiller KK (2010) Age-related loss of spiral ganglion neurons. *Hear Res* 264(1-2):93–97.

- Bedrick EJ, Hill JR (1990) Outlier tests for logistic regression: A conditional approach. *Biometrika* 77(4):815–827.
- Bellier L, Mazzuca M, Thai-Van H, Caclin A, Laboissière R (2013) Categorization of speech in early auditory evoked responses. *Conf Interspeech Lyon*, pp 837–841.
- Bellier L, Caclin A, Vesson J-F, Veuillet E, Thai-Van, H (2014) [Speech ABR: toward objective audiometry in the service of hearing aid?]. *[The Key Points in Audiology, Cerebral Plasticity and Rehabilitation of Deafness]*, eds Thai-Van H (L’Européenne d’Edition; Regimedia), pp 69–76.
- Bellier L, et al. (2015a) Topographic recordings of auditory evoked potentials to speech: Subcortical and cortical responses. *Psychophysiology* 52(4):594–599.
- Bellier L, et al. (2015b) Speech Auditory Brainstem Response through hearing aid stimulation. *Hear Res* 325:49–54.
- Bertoli S, Probst R, Bodmer D (2011) Late auditory evoked potentials in elderly long-term hearing-aid users with unilateral or bilateral fittings. *Hear Res* 280(1-2):58–69.
- Campbell T, Kerlin JR, Bishop CW, Miller LM (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear* 33(1):144–150.
- Carlsson P-I, Hall M, Lind K-J, Danermark B (2011) Quality of life, psychosocial consequences, and audiological rehabilitation after sudden sensorineural hearing loss. *Int J Audiol* 50(2):139–144.
- Ciorba A, Bianchini C, Pelucchi S, Pastore A (2012) The impact of hearing loss on the quality of life of elderly adults. *Clin Interv Aging* 7:159–163.
- Dajani HR, Heffernan BP, Giguere C (2013) Improving hearing aid fitting using the speech-evoked auditory brainstem response. *Conf Proc IEEE Eng Med Biol Soc* 2013:2812–2815.
- Dawes P, Munro KJ, Kalluri S, Edwards B (2013) Brainstem processing following unilateral and bilateral hearing-aid amplification. *Neuroreport* 24(6):271–275.
- Dawes P, Munro KJ, Kalluri S, Edwards B (2014) Auditory acclimatization and hearing aids: late auditory evoked potentials and speech recognition following unilateral and bilateral amplification. *J Acoust Soc Am* 135(6):3560–3569.
- Gabriel D, Veuillet E, Vesson JF, Collet L (2006) Rehabilitation plasticity: influence of hearing aid fitting on frequency discrimination performance near the hearing-loss cut-off. *Hear Res* 213(1-2):49–57.
- Gatehouse S (1992) The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. *J Acoust Soc Am* 92(3):1258–1268.
- Gatehouse S (1993) Role of perceptual acclimatization in the selection of frequency responses for hearing aids. *J Am Acad Audiol* 4(5):296–306.
- Gates GA, Mills JH (2005) Presbycusis. *Lancet* 366(9491):1111–1120.
- Johnson K (2003) *Acoustic and Auditory Phonetics* (Wiley).
- Kluk K, Moore BCJ (2006) Dead regions in the cochlea and enhancement of frequency discrimination: Effects of audiogram slope, unilateral versus bilateral loss, and hearing-aid use. *Hear Res* 222(1-2):1–15.

- Köbler S, Rosenhall U (2002) Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. *Int J Audiol* 41(7):395–400.
- Lavie L, Banai K, Karni A, Attias J (2015) Hearing Aid Induced Plasticity in the Auditory System of the Elderly: Evidence from Speech Perception. *J Speech Lang Hear Res*. doi:10.1044/2015_JSLHR-H-14-0225.
- Lefèvre F (2006) [Phonetic confusions induced by peripheral sensorineural hearing loss]. *SFA News*. Available at: <http://sfaudiologie.fr/Drupal/node/37> (in French).
- Liberman AM, Delattre PC, Cooper FS, Gerstman LJ (1954) The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol Monographs* 68(8):1–13.
- Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54(5):358–368.
- Lin FR, et al. (2011) Hearing loss and incident dementia. *Arch Neurol* 68(2):214–220.
- Munro KJ, Lutman ME (2003) The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. *J Acoust Soc Am* 114(1):484–495.
- Munro KJ, Trotter JH (2006) Preliminary evidence of asymmetry in uncomfortable loudness levels after unilateral hearing aid experience: evidence of functional plasticity in the adult auditory system. *Int J Audiol* 45(12):684–688.
- Munro KJ, Pisareva NY, Parker DJ, Purdy SC (2007) Asymmetry in the auditory brainstem response following experience of monaural amplification. *Neuroreport* 18(17):1871–1874.
- Munro KJ (2008) Reorganization of the adult auditory system: perceptual and physiological evidence from monaural fitting of hearing aids. *Trends Amplif* 12(3):254–271.
- Munro KJ, Merrett JF (2013) Brainstem plasticity and modified loudness following short-term use of hearing aids. *J Acoust Soc Am* 133(1):343–349.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97–113.
- Ouda L, Profant O, Syka J (2015) Age-related changes in the central auditory system. *Cell Tissue Res* 361(1):337–358.
- Panza F, Solfrizzi V, Logroscino G (2015) Age-related hearing impairment-a risk factor and frailty marker for dementia and AD. *Nat Rev Neurol* 11(3):166–175.
- Philibert B, Collet L, Vesson J-F, Veuillet E (2002) Intensity-related performances are modified by long-term hearing aid use: a functional plasticity? *Hear Res* 165(1-2):142–151.
- Philibert B, Collet L, Vesson JF, Veuillet E (2003) Auditory rehabilitation effects on speech lateralization in hearing-impaired listeners. *Acta Otolaryngol* 123(2):172–175.
- Philibert B, Collet L, Vesson J-F, Veuillet E (2005) The auditory acclimatization effect in sensorineural hearing-impaired listeners: evidence for functional plasticity. *Hear Res* 205(1-2):131–142.

- Plyler PN, Ananthanarayan AK (2001) Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. *J Am Acad Audiol* 12(10):523–533.
- Pregibon D (1981) Logistic Regression Diagnostics. *Ann Statist* 9(4):705–724.
- Robinson K, Gatehouse S (1995) Changes in intensity discrimination following monaural long-term use of a hearing aid. *J Acoust Soc Am* 97(2):1183–1190.
- Roth TN (2015) Aging of the auditory system. *Handb Clin Neurol* 129:357–373.
- Sakaguchi H, Tokita J, Müller U, Kachar B (2009) Tip links in hair cells: molecular composition and role in hearing loss. *Curr Opin Otolaryngol Head Neck Surg* 17(5):388–393.
- Simon C, Fourcin AJ (1978) Cross-language study of speech-pattern learning. *The Journal of the Acoustical Society of America* 63(3):925–935.
- Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324.
- Stapells DR, Oates P (1997) Estimation of the pure-tone audiogram by the auditory brainstem response: a review. *Audiol Neurotol* 2(5):257–280.
- Thai-Van H, Veuillet E, Norena A, Guiraud J, Collet L (2010) Plasticity of tonotopic maps in humans: influence of hearing loss, hearing aids and cochlear implants. *Acta Otolaryngol* 130(3):333–337.
- Warner-Czyz AD, Loy BA, Evans C, Wetsel A, Tobey EA (2015) Self-esteem in children and adolescents with hearing loss. *Trends Hear* 19. doi:10.1177/2331216515572615.
- Wilson RS, et al. (2007) Loneliness and risk of Alzheimer disease. *Arch Gen Psychiatry* 64(2):234–240.
- Wong ACY, Ryan AF (2015) Mechanisms of sensorineural cell damage, death and survival in the cochlea. *Front Aging Neurosci* 7:58.
- Yamasoba T, et al. (2013) Current concepts in age-related hearing loss: epidemiology and mechanistic pathways. *Hear Res* 303:30–38.

Troisième partie

Discussion générale

6. Discussion générale

6.1. Résumé des résultats

En suivant l'axe des générateurs neuronaux de la Speech ABR, nous avons d'abord mis au point une technique jusqu'alors inédite d'enregistrement topographique simultané de cette réponse et des PEA corticaux. Nous avons par la même occasion présenté pour la première fois une analyse de débruitage multi-électrodes ainsi rendue possible (Denoising Source Separation ; de Cheveigné & Simon, 2008), et montré la faisabilité de l'acquisition de la Speech ABR avec un système EEG doté d'un taux d'échantillonnage relativement bas (5 kHz, contre le minimum de 10 kHz à l'usage). La poursuite de notre investigation par l'étude des PEA intracrâniens a révélé l'existence d'une activité de type FFR dans le cortex auditif primaire humain, capable de suivre jusqu'à 200 Hz la fréquence fondamentale des *stimuli* de parole utilisés. Ce résultat non seulement excède la limite de 114 Hz observée précédemment (Steinschneider *et al.*, 2013), mais permet aussi de considérer la possibilité d'un encodage de la fréquence fondamentale par verrouillage de phase dans le cortex auditif primaire. Enfin, l'analyse comparée des topographies de scalp pour la Speech ABR et l'onde P50 des PEA corticaux a suggéré la participation de la FFR corticale sus-citée dans la génération de la réponse de scalp. Ainsi, nos résultats permettent de questionner l'interprétation de dizaines d'articles de la littérature se basant sur l'hypothèse d'une génération purement sous-corticale, qui déduisent de l'observation d'une altération de l'encodage de F0 la preuve d'une plasticité du tronc cérébral auditif.

Du côté de l'axe de la Speech ABR chez le malentendant appareillé, nous avons apporté à la littérature les tout premiers enregistrements de cette réponse avec stimulation directe *via* les aides auditives. Cette technique ouvre ainsi une ligne de recherche qui pourra bénéficier à de nombreux malentendants. De plus, l'étude longitudinale chez le malentendant appareillé montre une amélioration comportementale associée à une plasticité neuronale après un temps d'utilisation des aides auditives plus court que celui décrit jusqu'alors. En objectivant les bénéfices de l'appareillage auditif grâce à l'utilisation innovante de la Speech ABR, nous faisons un premier pas vers une audiométrie vocale objective ainsi que dans le sens d'une prise en charge optimisée des patients malentendants.

En résumé, les résultats issus des travaux de la présente thèse apportent une contribution significative à la littérature de la Speech ABR, à la fois en amenant à une reconsidération de sa nature purement sous-corticale, en offrant de nouvelles techniques de stimulation, d'enregistrement et d'analyses à la méthodologie existante, et en objectivant de façon originale la plasticité neuronale induite par l'utilisation d'aides auditives chez le malentendant presbyacousique.

6.2. Discussion, limites et perspectives

6.2.1. Méthodologie de la Speech ABR

L'un des premiers résultats de cette thèse est la possibilité d'enregistrer la Speech ABR à l'aide d'un système EEG 32 voies doté d'une « faible » fréquence d'échantillonnage (relativement aux autres études du domaine). Dans l'étude 1, nous avons notamment montré que le signal acquis à 5 kHz (avec filtre anti-repliement à 1 kHz) était fortement corrélé et de qualité comparable au signal acquis à 13,3 kHz à l'aide d'un second système EEG 4 voies. En accord avec le théorème d'échantillonnage de Nyquist-Shannon, cette première fréquence suffit en effet à enregistrer l'intégralité de l'information contenue dans la FFR d'enveloppe, puisque la fréquence de coupure du tronc cérébral supérieur est généralement admise aux alentours de 1,5 kHz (et celle du cortex auditif primaire à une valeur encore inférieure). Deux limites existent cependant : le signal bénéficie d'une moindre résolution temporelle, ce qui peut réduire la précision des mesures de latence (Skoe & Kraus, 2010a) ; aussi, cette faible fréquence d'échantillonnage doublée du filtre anti-repliement ne permet pas d'étudier pleinement la FFR spectrale, particulièrement pour le suivi de la trajectoire du second formant (Plyler & Ananthanaryan, 2001). L'évolution technologique constante des amplificateurs EEG devrait permettre dans quelques années de contourner cette limite, en associant haute fréquence d'échantillonnage, faible niveau de bruit et enregistrement sur un grand nombre de voies, en rendant ainsi possible une acquisition quasi exhaustive de l'information électrophysiologique générée en réponse aux sons de parole.

L'acquisition topographique de la Speech ABR rend possible l'analyse des topographies pour éventuellement confirmer les générateurs des différentes composantes (FFR d'enveloppe et spectrale, réponses transitoires) ou suivre leurs altérations dans différentes expertises ou déficiences auditives, voire même dissocier les activités des différentes sources pour effectuer des analyses de connectivité entre celles-ci. Aussi, les données acquises par cette technique peuvent bénéficier d'algorithmes de débruitage pour alors réduire le temps d'enregistrement requis (Bharadwaj & Shinn-Cunningham, 2014), ce qui permettrait d'étudier un plus grand nombre de syllabes ou une plus grande variété de *stimuli* lors d'une même session.

A notre connaissance, aucune Speech ABR n'a en effet été montrée pour des consonnes fricatives, spirantes ou latérales, et seulement très rarement pour des nasales. Or, il est tout à fait possible d'imaginer des effets spécifiques à ces consonnes et qui ne s'observeraient pas pour les occlusives classiquement utilisées ; par exemple, nous pourrions supposer un encodage dans le bruit d'autant plus dégradé pour des fricatives qui y sont particulièrement sensibles, que pour des occlusives relativement robustes du fait de leur énergie plus importante liée à l'explosion du relâchement du flux d'air.

Concernant la stimulation *via* les aides auditives, trois variantes peuvent être notées : transmission du signal par des haut-parleurs en champ libre puis amplification par les appareils auditifs, ou transmission directe du signal aux appareils par voie sans-fil ou filaire.

Nous n'avons pas exploré la première, du fait des changements de position et mouvements de tête du sujet, susceptibles de déphaser les essais unitaires et donc de dégrader la Speech ABR. Aussi, cette première variante n'est pas aisément compatible avec la condition contrôle pour l'artéfact de stimulation, consistant à délivrer le *stimulus* en ayant préalablement bouché les écouteurs et placé ceux-ci hors des conduits auditifs : il faudrait alors boucher les oreilles du sujet, sans certitude qu'il ne reçoive pas le signal par vibration osseuse suivant l'intensité de stimulation des haut-parleurs. Nous avons au contraire longuement exploré les deux variantes « directes », en obtenant une Speech ABR de qualité dans les deux cas. Néanmoins, la fiabilité temporelle varie en fonction des marques et des modèles d'appareils auditifs, aussi cette méthodologie de stimulation requiert-elle un contrôle très fin et l'enregistrement de sujets pilote avant tout début de protocole.

Avec l'évolution de l'audioprothèse vers une plus grande précision dans l'évaluation des troubles auditifs, ces difficultés techniques pourraient être prises en charge directement par les fabricants, avec l'implémentation d'une voie contrôlée de transmission directe du signal aux prothèses. Plus généralement, la vérification de la sortie acoustique des transducteurs utilisés est primordiale dans l'étude de la Speech ABR : le signal est bien souvent distordu, amoindrissant de fait la pertinence des analyses de corrélation croisée entre *stimulus* et réponse électrophysiologique.

Rappelons enfin que la Speech ABR a été annoncée comme une mesure idéale pour la prise en charge audioprothétique des troubles auditifs, étant donné son encodage bas-niveau des sons de parole testés (Anderson & Kraus, 2013 ; Dajani *et al.*, 2013). L'étude 3 de cette thèse constitue donc un précédent dans cette application promise à un riche avenir, au bénéfice premier du patient malentendant.

6.2.2. Générateurs neuronaux de la Speech ABR

L'existence de la FFR corticale telle que nous l'avons enregistrée en SEEG ne semble pas sujette à contestation : observée uniquement dans le cortex auditif primaire, avec l'utilisation de la polarité alternée et du référencement bipolaire, il ne peut s'agir ni d'un artéfact, ni d'une activité provenant du tronc cérébral supérieur. Celle-ci n'exclut absolument pas l'existence de la FFR sous-corticale, et ne concerne d'ailleurs peut-être que le suivi de la fréquence fondamentale.

La participation de la FFR corticale à la Speech ABR enregistrée au niveau du scalp est cependant plus discutable, bien qu'il serait étonnant que l'on parvienne à détecter l'activité d'une source profonde (lemnisque latéral ou colliculus inférieur) depuis le scalp sans pour autant détecter la FFR corticale. Cela pourrait toutefois être rendu possible par une architecture corticale inadéquate à la propagation du signal vers le scalp. L'organisation du cortex auditif primaire en colonnes corticales rend cependant peu crédible cette explication potentielle (Galaburda & Sanides, 1980). Pour aller plus loin dans la vérification de cette participation corticale suggérée, nous avons effectué plusieurs reconstructions de sources des topographies présentées dans l'étude 2, mais l'absence d'information sur l'anatomie des sujets et la localisation des électrodes n'a pas permis à ces tentatives d'aboutir. De futures études sont donc nécessaires pour confirmer notre observation, et ainsi pouvoir asseoir plus solidement la nécessité d'une réinterprétation des résultats de la littérature *Speech ABR*. Notons qu'un tel exercice n'affecterait en rien l'intérêt de la Speech ABR, qui resterait un marqueur privilégié de l'encodage des sons de parole.

La question de la spécificité aux sons de parole de la FFR corticale mesurée dans l'étude 2 se pose de la même manière qu'elle a été posée par le passé pour la Speech ABR. Nous avons pour cela acquis des données de SEEG en réponse à des sons de complexité graduelle (tons purs, tons complexes, parole synthétique, parole naturelle), que nous analyserons en détail dans les prochains mois. Une première plongée dans les données suggère une activité plus ample en réponse à la parole naturelle, par rapport aux trois autres catégories de *stimuli*, dans le cortex auditif primaire. Une telle observation va dans le sens des résultats d'IRMf sur les aires corticales spécifiques à la voix (Ethofer *et al.*, 2009 ; Leaver & Rauschecker, 2010), en bénéficiant ici des avantages de la SEEG pour étudier plus précisément les différences spectro-temporelles fines de l'encodage de ces différents sons.

Enfin, la Speech ABR reflète l'encodage spectro-temporel bas-niveau des *stimuli* de parole testés, et en admettant que cette représentation soit maintenue jusqu'au cortex auditif primaire, il serait intéressant de suivre cette information en aval, c'est-à-dire dans les aires corticales responsables d'en extraire les phonèmes, par exemple. Ainsi, plutôt que les associations « *stimulus* - réponse comportementale », « *stimulus* - encodage neuronal bas-niveau », voire « *stimulus* - encodage neuronal bas-niveau - réponse comportementale » étudiées respectivement par la psychoacoustique, la Speech ABR seule et la Speech ABR associée à une tâche comportementale, nous pourrions observer le lien entre encodage neuronal bas-niveau, variable et majoritairement extrinsèque, et encodage neuronal haut-niveau, abstrait et majoritairement intrinsèque. Cette approche offrirait davantage d'intérêt dans la compréhension des troubles de l'audition, par exemple en estimant le degré de dégradation de

l'encodage neuronal à partir duquel les aires phonologiques ne pourraient plus « lire » l'information linguistique.

6.2.3. Plasticité neuronale chez le malentendant appareillé

Les résultats préliminaires de l'étude 4 mettent en évidence des améliorations comportementales et électrophysiologiques induites par l'utilisation d'aides auditives. Nous observons notamment une meilleure identification du trait phonétique de nasalité, une diminution globale de l'encodage de F0 anormalement élevé auparavant, et une diminution pour une syllabe nasale de l'amplitude du complexe cortical N1r-P2r, également anormalement élevé par rapport aux sujets témoins avant le début de la remédiation auditive. Si une amélioration comportementale moins franche avait déjà été observée pour la même durée d'utilisation des appareils auditifs (Philibert *et al.*, 2003), les variations électrophysiologiques n'avaient été retrouvées qu'après des périodes plus longues, se comptant parfois en années (Bertoli *et al.*, 2011). Nous apportons ainsi pour la première fois une preuve objective des effets bénéfiques de l'appareillage auditif en un si court laps de temps. De nombreuses analyses restent toutefois à faire pour approfondir ce résultat.

C'est aussi la première fois qu'une étude de Speech ABR est réalisée chez le malentendant appareillé. De rares travaux précédents avaient en effet décrit l'impact de pertes auditives sur la Speech ABR, cependant chez le patient malentendant non-appareillé, et avec une stimulation *via* écouteurs de type *inserts* (Anderson *et al.*, 2013a ; Parbery-Clark *et al.*, 2013b). Grâce à notre méthodologie innovante de stimulation directe *via* les aides auditives, nous entamons donc une ligne de recherche qui pourra mener à terme à une nette amélioration de la prise en charge des patients malentendants.

Enfin, l'utilisation des aides auditives seules ne suffit pas en soi à regagner un maximum d'intelligibilité de la parole dans le bruit, par exemple : elle apporte des informations hautes-fréquences jusque-là en partie perdues, mais c'est la confrontation aux sons de l'environnement qui va permettre au patient de ré-étiqueter le monde autour de lui, en associant des sons alors nouveaux aux objets connus. Les entraînements auditifs potentialisent cette démarche cognitive plus que sensorielle, lesquels sont également prometteurs pour potentialiser l'effet des aides auditives. La combinaison de ces deux approches, nommément audioprothétique et cognitive, améliorée en de multiples aspects par l'utilisation de la Speech ABR, pourrait alors constituer une stratégie optimale pour la remédiation des troubles de l'audition.

Références bibliographiques

- Abrams DA, Nicol T, Zecker SG, Kraus N (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. *J Neurosci* 26(43):11131–11137.
- Abrams DA, Nicol T, Zecker S, Kraus N (2010) Rapid acoustic processing in the auditory brainstem is not related to cortical asymmetry for the syllable rate of speech. *Clin Neurophysiol* 121(8):1343–1350.
- Achor LJ, Starr A (1980a) Auditory brain stem responses in the cat. I. Intracranial and extracranial recordings. *Electroencephalogr Clin Neurophysiol* 48(2):154–173.
- Achor LJ, Starr A (1980b) Auditory brain stem responses in the cat. II. Effects of lesions. *Electroencephalogr Clin Neurophysiol* 48(2):174–190.
- Ahadi M, Pourbakht A, Jafari AH, Shirjian Z, Jafarpisheh AS (2014a) Gender disparity in subcortical encoding of binaurally presented speech stimuli: an auditory evoked potentials study. *Auris Nasus Larynx* 41(3):239–243.
- Ahadi M, Pourbakht A, Jafari AH, Jalaie S (2014b) Effects of stimulus presentation mode and subcortical laterality in speech-evoked auditory brainstem responses. *Int J Audiol* 53(4):243–249.
- Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245(1-2):35–47.
- Aiken SJ, Purcell D (2013) Sensitivity to stimulus polarity in speech-evoked frequency-following responses. *ICA 2013 Montreal* (ASA, Montreal, Canada), p 050121.
- Akhoun I, Gallégo S, Moulin A, Ménard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H (2008a) The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 119(4):922–933.
- Akhoun I, Moulin A, Jeanvoine A, Ménard M, Buret F, Vollaie C, Scorretti R, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H (2008b) Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. *J Neurosci Methods* 175(2):196–205.
- Akhoun I, Berger-Vachon C, Collet L (2010) Approche empirique pour déterminer les générateurs neurophysiologiques sous-jacents des potentiels évoqués auditifs engendrés par des sons de parole (Speech ABR). *IRBM* 31(4):242–256.
- Albouy P, Mattout J, Bouet R, Maby E, Sanchez G, Aguera P-E, Daligault S, Delpuech C, Bertrand O, Caclin A, Tillmann B (2013) Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* 136(Pt 5):1639–1661.
- Anderson S, Kraus N (2010a) Objective neural indices of speech-in-noise perception. *Trends Amplif* 14(2):73–83.
- Anderson S, Kraus N (2010b) Sensory-cognitive interaction in the neural encoding of speech in noise: a review. *J Am Acad Audiol* 21(9):575–585.
- Anderson S, Skoe E, Chandrasekaran B, Kraus N (2010a) Neural timing is linked to speech perception in noise. *J Neurosci* 30(14):4922–4926.

- Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N (2010b) Brainstem correlates of speech-in-noise perception in children. *Hear Res* 270(1-2):151–157.
- Anderson S, Parbery-Clark A, Yi H-G, Kraus N (2011) A neural basis of speech-in-noise perception in older adults. *Ear Hear* 32(6):750–757.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. *J Neurosci* 32(41):14156–14164.
- Anderson S, Kraus N (2013) The Potential Role of the cABR in Assessment and Management of Hearing Impairment. *Int J Otolaryngol* 2013:604729.
- Anderson S, Parbery-Clark A, White-Schwoch T, Drehobl S, Kraus N (2013a) Effects of hearing loss on the subcortical representation of speech cues. *J Acoust Soc Am* 133(5):3030–3038.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2013b) Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *J Speech Lang Hear Res* 56(1):31–43.
- Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N (2013c) A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear Res* 300:18–32.
- Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N (2013d) Reversal of age-related neural timing delays with training. *Proc Natl Acad Sci USA* 110(11):4357–4362.
- Anderson S, White-Schwoch T, Choi HJ, Kraus N (2014) Partial maintenance of auditory-based cognitive training benefits in older adults. *Neuropsychologia* 62:286–296.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2015) Development of subcortical speech representation in human infants. *J Acoust Soc Am* 137(6):3346.
- Baddeley A, Gathercole S, Papagno C (1998) The phonological loop as a language learning device. *Psychol Rev* 105(1):158–173.
- Banai K, Nicol T, Zecker SG, Kraus N (2005) Brainstem timing: implications for cortical processing and literacy. *J Neurosci* 25(43):9850–9857.
- Banai K, Abrams D, Kraus N (2007) Sensory-based learning disability: Insights from brainstem processing of speech sounds. *Int J Audiol* 46(9):524–532.
- Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009) Reading and subcortical auditory function. *Cereb Cortex* 19(11):2699–2707.
- Basu M, Krishnan A, Weber-Fox C (2010) Brainstem correlates of temporal auditory processing in children with specific language impairment. *Dev Sci* 13(1):77–91.
- Bear MF, Connors BW, Paradiso MA (2007) *Neurosciences: à la découverte du cerveau* (Pradel, Wolters Kluwer France).
- Bharadwaj HM, Shinn-Cunningham BG (2014) Rapid acquisition of auditory subcortical steady state responses using multichannel recordings. *Clin Neurophysiol* 125(9):1878–1888.
- Bidelman GM, Krishnan A (2009) Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J Neurosci* 29(42):13165–13171.
- Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.

- Bidelman GM, Krishnan A (2011) Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport* 22(5):212–216.
- Bidelman GM, Gandour JT, Krishnan A (2011a) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 23(2):425–434.
- Bidelman GM, Gandour JT, Krishnan A (2011b) Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn* 77(1):1–10.
- Bidelman GM, Gandour JT, Krishnan A (2011c) Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neurosci Lett* 503(3):203–207.
- Bidelman GM, Krishnan A, Gandour JT (2011d) Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci* 33(3):530–538.
- Bidelman GM (2013) The role of the auditory brainstem in processing musically relevant pitch. *Front Psychol* 4:264.
- Bidelman GM, Hutka S, Moreno S (2013a) Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8(4):e60676.
- Bidelman GM, Moreno S, Alain C (2013b) Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79:201–212.
- Bidelman GM (2014) Objective information-theoretic algorithm for detecting brainstem-evoked responses to complex stimuli. *J Am Acad Audiol* 25(8):715–726.
- Bidelman GM, Villafuerte JW, Moreno S, Alain C (2014a) Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol Aging* 35(11):2526–2540.
- Bidelman GM, Weiss MW, Moreno S, Alain C (2014b) Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci* 40(4):2662–2673.
- Bidelman GM, Alain C (2015) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci* 35(3):1240–1249.
- Bidelman GM (2015a) Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J Neurosci Methods*. doi:10.1016/j.jneumeth.2014.12.019.
- Bidelman GM (2015b) Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hear Res* 323:68–80.
- Bones O, Hopkins K, Krishnan A, Plack CJ (2014) Phase locked neural activity in the human brainstem predicts preference for musical consonance. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2014.03.011.
- Borg E, Zakrisson JE (1975) The activity of the stapedius muscle in man during vocalization. *Acta Otolaryngol* 79(5-6):325–333.
- Buchwald JS, Huang C (1975) Far-field acoustic response: origins in the cat. *Science* 189(4200):382–384.
- Campbell T, Kerlin JR, Bishop CW, Miller LM (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear* 33(1):144–150.

- Chabout J, Serreau P, Ey E, Bellier L, Aubin T, Bourgeron T, Granon S (2012) Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS ONE* 7(1):e29401.
- Chandrasekaran B, Hornickel J, Skoe E, Nicol T, Kraus N (2009) Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64(3):311–319.
- Chandrasekaran B, Kraus N (2010a) Music, Noise-Exclusion, and Learning. *Music Percept* 27(4):297–306.
- Chandrasekaran B, Kraus N (2010b) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47(2):236–246.
- Chandrasekaran B, Kraus N, Wong PCM (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. *J Neurophysiol* 107(5):1325–1336.
- Check Hayden E (2015) Anti-ageing pill pushed as bona fide drug. *Nature News* 522(7556):265.
- Choi JM, Purcell DW, Coyne J-AM, Aiken SJ (2013) Envelope following responses elicited by english sentences. *Ear Hear* 34(5):637–650.
- Chomsky N, Halle M (1968) *The Sound Pattern of English* (MIT Press).
- Clinard CG, Tremblay KL, Krishnan AR (2010) Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hear Res* 264(1-2):48–55.
- Clinard CG, Tremblay KL (2013) Aging degrades the neural encoding of simple and complex sounds in the human brainstem. *J Am Acad Audiol* 24(7):590–599.
- Clinard CG, Cotter CM (2015) Neural representation of dynamic frequency is degraded in older adults. *Hear Res* 323:91–98.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 112(5):758–767.
- Dajani HR, Heffernan BP, Giguere C (2013) Improving hearing aid fitting using the speech-evoked auditory brainstem response. *Conf Proc IEEE Eng Med Biol Soc* 2013:2812–2815.
- Dau T, Wegner O, Mellert V, Kollmeier B (2000) Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. *J Acoust Soc Am* 107(3):1530–1540.
- Davies P, Gregersen NH (2014) *Information and the Nature of Reality: From Physics to Metaphysics* (Cambridge University Press).
- Davis H, Derbyshire A, Lurie M, Saul L (1934) The electric response of the cochlea. *American Journal of Physiology–Legacy Content* 107(2):311–332.
- de Cheveigné A, Simon JZ (2008) Denoising based on spatial filtering. *J Neurosci Methods* 171(2):331–339.
- Debray R (1991) *Cours de médiologie générale* (Gallimard).
- Dhar S, et al. (2009) Exploring the relationship between physiological measures of cochlear and brainstem function. *Clin Neurophysiol* 120(5):959–966.
- Dobie RA, Berlin CI (1979) Binaural interaction in brainstem-evoked responses. *Arch Otolaryngol* 105(7):391–398.

- Eggermont JJ (1974) Basic principles for electrocochleography. *Acta Otolaryngol Suppl* 316:7–16.
- Eggermont JJ, Odenthal DW (1974) Methods in electrocochleography. *Acta Otolaryngol Suppl* 316:17–24.
- Elkabariti RH, Khalil LH, Husein R, Talaat HS (2014) Speech evoked auditory brainstem response findings in children with epilepsy. *Int J Pediatr Otorhinolaryngol* 78(8):1277–1280.
- Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of emotional information in voice-sensitive cortices. *Curr Biol* 19(12):1028–1033.
- Fant G (1960) *Acoustic Theory of Speech Production* (Mouton, The Hague).
- Filippini R, Befi-Lopes DM, Schochat E (2012) Efficacy of auditory training using the auditory brainstem response to complex sounds: auditory processing disorder and specific language impairment. *Folia Phoniatr Logop* 64(5):217–226.
- Fu Q, Liang Y, Zou A, Wang T (2015) [Relationships of electrophysiological characteristic between speech evoked auditory brainstem response and auditory mismatch negativity]. *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 29(1):39–44.
- Fujihira H, Shiraishi K (2015) Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clin Neurophysiol* 126(1):96–102.
- Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190(3):597–610.
- Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci USA* 78(4):2643–2647.
- Galbraith GC, Arbagey PW, Branski R, Comerici N, Rector PM (1995) Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6(17):2363–2367.
- Galbraith GC, Jhaveri SP, Kuo J (1997) Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. *Electroencephalogr Clin Neurophysiol* 102(1):46–53.
- Galbraith GC, Bhuta SM, Choate AK, Kitahara JM, Mullen TA Jr (1998) Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9(8):1889–1893.
- Galbraith GC, Bagasan B, Sulahian J (2001) Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Percept Mot Skills* 92(1):99–106.
- Galbraith GC, Amaya EM, de Rivera JMD, Donan NM, Duong MT, Hsu JN, Tran K, Tsang LP (2004) Brain stem evoked response to forward and reversed speech in humans. *Neuroreport* 15(13):2057–2060.
- Gardi J, Merzenich M, McKean C (1979) Origins of the scalp recorded frequency-following response in the cat. *Audiology* 18(5):358–381.
- Garnero L (2001) Les bases physiques et physiologiques de la Magnétoencéphalographie et de l'Electroencéphalographie (CNRS-UPR640-LENA).
- Giard MH, Collet L, Bouchet P, Pernier J (1994) Auditory selective attention in the human cochlea. *Brain Res* 633(1-2):353–356.
- Gilbert AL, Regier T, Kay P, Ivry RB (2006) Whorf hypothesis is supported in the right visual field but not the left. *Proc Natl Acad Sci USA* 103(2):489–494.

- Gnanateja GN, Ranjan R, Maruthy S (2012) Physiological basis of the encoding of speech evoked frequency following responses. *J Ind Inst Speech Hear* 31:215–219.
- Hashimoto I, Ishiyama Y, Yoshimoto T, Nemoto S (1981) Brain-stem auditory-evoked potentials recorded directly from human brain-stem and thalamus. *Brain* 104(Pt 4):841–859.
- Hayes EA, Tiippana K, Nicol TG, Sams M, Kraus N (2003a) Integration of heard and seen speech: a factor in learning disabilities in children. *Neurosci Lett* 351(1):46–50.
- Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N (2003b) Neural plasticity following auditory training in children with learning problems. *Clin Neurophysiol* 114(4):673–684.
- Herdman AT, Stapells DK (2003) Auditory steady-state response thresholds of adults with sensorineural hearing impairments. *Int J Audiol* 42(5):237–248.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.
- Hofman PM, Van Riswick JG, Van Opstal AJ (1998) Relearning sound localization with new ears. *Nat Neurosci* 1(5):417–421.
- Hoijer H (1954) The Sapir-Whorf Hypothesis. *Language in Culture* (University of Chicago Press), pp 92–105.
- Hornickel J, Skoe E, Kraus N (2009a) Subcortical laterality of speech encoding. *Audiol Neurotol* 14(3):198–207.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009b) Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA* 106(31):13022–13027.
- Hornickel J, Chandrasekaran B, Zecker S, Kraus N (2011) Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. *Behav Brain Res* 216(2):597–605.
- Hornickel J, Anderson S, Skoe E, Yi H-G, Kraus N (2012a) Subcortical representation of speech fine structure relates to reading ability. *Neuroreport* 23(1):6–9.
- Hornickel J, Knowles E, Kraus N (2012b) Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hear Res* 284(1-2):52–58.
- Hornickel J, Zecker SG, Bradlow AR, Kraus N (2012c) Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc Natl Acad Sci USA* 109(41):16731–16736.
- Hornickel J, Kraus N (2013) Unstable representation of sound: a biological marker of dyslexia. *J Neurosci* 33(8):3500–3504.
- Hornickel J, Lin D, Kraus N (2013) Speech-evoked auditory brainstem responses reflect familial and cognitive influences. *Dev Sci* 16(1):101–110.
- Hudspeth AJ (2014) Integrating the active process of hair cells with cochlear function. *Nat Rev Neurosci* 15(9):600–614.
- IPA Chart, <http://www.internationalphoneticassociation.org/content/ipa-chart>, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2005 International Phonetic Association.
- Jafari Z, Malayeri S (2014) Effects of congenital blindness on the subcortical representation of speech cues. *Neuroscience* 258:401–409.
- Jafari Z, Malayeri S, Rostami R (2015) Subcortical encoding of speech cues in children with attention deficit hyperactivity disorder. *Clin Neurophysiol* 126(2):325–332.

- Jakobson R, Fant G, Halle M (1951) Preliminaries to Speech Analysis. The distinctive features and their correlates (MIT Press, Cambridge MA).
- Jakobson R, Halle M (1956) Fundamentals of Language (Walter de Gruyter).
- Jeng F-C, Schnabel E, Dickman B, Hu J, Li X, Lin CD, Chung HK (2010) Early maturation of frequency-following responses to voice pitch in infants with normal hearing. *Percept Mot Skills* 111(3):765–784.
- Jeng F-C, Chung H-K, Lin C-D, Dickman B, Hu J (2011a) Exponential modeling of human frequency-following responses to voice pitch. *Int J Audiol* 50(9):582–593.
- Jeng F-C, Costilow CE, Stangherlin DP, Lin C-D (2011b) Relative power of harmonics in human frequency-following responses associated with voice pitch in American and Chinese adults. *Percept Mot Skills* 113(1):67–86.
- Jeng F-C, Hu J, Dickman B, Lin C-Y, Lin C-D, Wang C-Y, Chung H-K, Li X (2011c) Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. *Int J Audiol* 50(1):14–26.
- Jeng F-C, Hu J, Dickman B, Montgomery-Reagan K, Tong M, Wu G, Lin C-D (2011d) Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear Hear* 32(6):699–707.
- Jeng F-C, Peris KS, Hu J, Lin C-D (2013) Evaluation of an automated procedure for detecting frequency-following responses in American and Chinese neonates. *Percept Mot Skills* 116(2):456–465.
- Jewett DL, Romano MN, Williston JS (1970) Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 167(3924):1517–1518.
- Jewett DL, Williston JS (1971) Auditory-evoked far fields averaged from the scalp of humans. *Brain* 94(4):681–696.
- Johnson KL, Nicol TG, Kraus N (2005) Brain stem response to speech: a biological marker of auditory processing. *Ear Hear* 26(5):424–434.
- Johnson KL, Nicol TG, Zecker SG, Kraus N (2007) Auditory brainstem correlates of perceptual timing deficits. *J Cogn Neurosci* 19(3):376–385.
- Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N (2008a) Brainstem encoding of voiced consonant-vowel stop syllables. *Clin Neurophysiol* 119(11):2623–2635.
- Johnson KL, Nicol T, Zecker SG, Kraus N (2008b) Developmental plasticity in the human auditory brainstem. *J Neurosci* 28(15):4000–4007.
- King C, Warrier CM, Hayes E, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett* 319(2):111–115.
- Kochkin S (2010) MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal* 63(1):19–20.
- Kouni SN, Papadeas ES, Varakis IN, Kouvelas HD, Koutsojannis CM (2006) Auditory brainstem responses in dyslexia: comparison between acoustic click and verbal stimulus events. *J Otolaryngol* 35(5):305–309.
- Kouni SN, Giannopoulos S, Ziavra N, Koutsojannis C (2013) Brainstem auditory evoked potentials with the use of acoustic clicks and complex verbal sounds in young adults with learning disabilities. *Am J Otolaryngol*. doi:10.1016/j.amjoto.2013.07.004.

- Kouni SN, Koutsojannis C, Ziavra N, Giannopoulos S (2014) A novel method of brainstem auditory evoked potentials using complex verbal stimuli. *N Am J Med Sci* 6(8):418–421.
- Kraus N (2001) Auditory pathway encoding and neural plasticity in children with learning problems. *Audiol Neurotol* 6(4):221–227.
- Kraus N, Skoe E, Parbery-Clark A, Ashley R (2009) Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann N Y Acad Sci* 1169:543–557.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11(8):599–605.
- Kraus N (2012) Biological impact of music and software-based auditory training. *J Commun Disord* 45(6):403–410.
- Kraus N, Anderson S (2014) Bigger is not better: effects of hearing loss on central processing. *Hear J* 67(1):24.
- Krishnan A (1999) Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiol Neurotol* 4(2):95–103.
- Krishnan A, Parkinson J (2000) Human frequency-following response: representation of tonal sweeps. *Audiol Neurotol* 5(6):312–321.
- Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166(1-2):192–201.
- Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189(1-2):1–12.
- Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25(1):161–168.
- Krishnan A, Gandour JT (2009) The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang* 110(3):135–148.
- Krishnan A, Gandour JT, Bidelman GM, Swaminathan J (2009a) Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport* 20(4):408–413.
- Krishnan A, Swaminathan J, Gandour JT (2009b) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 21(6):1092–1105.
- Krishnan A, Agrawal S (2010) Human frequency-following response to speech-like sounds: correlates of off-frequency masking. *Audiol Neurotol* 15(4):221–228.
- Krishnan A, Gandour JT, Smalt CJ, Bidelman GM (2010a) Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. *Brain Lang* 114(3):193–198.
- Krishnan A, Gandour JT, Bidelman GM (2010b) The effects of tone language experience on pitch processing in the brainstem. *J Neurolinguistics* 23(1):81–95.
- Krishnan A, Gandour JT, Bidelman GM (2010c) Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Res* 1313:124–133.
- Krishnan A, Plack CJ (2011) Neural encoding in the human brainstem relevant to the pitch of complex tones. *Hear Res* 275(1-2):110–119.

- Krishnan A, Gandour JT, Ananthakrishnan S, Bidelman GM, Smalt CJ (2011a) Functional ear (a)symmetry in brainstem neural activity relevant to encoding of voice pitch: a precursor for hemispheric specialization? *Brain Lang* 119(3):226–231.
- Krishnan A, Gandour JT, Ananthakrishnan S, Bidelman GM, Smalt CJ (2011b) Linguistic status of timbre influences pitch encoding in the brainstem. *Neuroreport* 22(16):801–803.
- Krishnan A, Gandour JT, Bidelman GM (2012) Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. *Neuroreport* 23(8):498–502.
- Krizman JL, Skoe E, Kraus N (2010) Stimulus rate and subcortical auditory processing of speech. *Audiol Neurotol* 15(5):332–342.
- Krizman J, Marian V, Shook A, Skoe E, Kraus N (2012a) Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc Natl Acad Sci USA* 109(20):7877–7881.
- Krizman J, Skoe E, Kraus N (2012b) Sex differences in auditory subcortical function. *Clin Neurophysiol* 123(3):590–597.
- Krizman J, Skoe E, Marian V, Kraus N (2014) Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain Lang* 128(1):34–40.
- Krizman J, et al. (2015) Continued maturation of auditory brainstem function during adolescence: A longitudinal approach. *Clin Neurophysiol*. doi:10.1016/j.clinph.2015.01.026.
- Laborit H (1985) *La nouvelle grille* (Gallimard).
- Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8(4):194–208.
- Larkin RP, Sutherland PJ (1977) Migrating birds respond to Project Seafarer's electromagnetic field. *Science* 195(4280):777–779.
- Leask MJ (1977) A physicochemical mechanism for magnetic field detection by migratory birds and homing pigeons. *Nature* 267(5607):144–145.
- Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30(22):7604–7612.
- Lee KM, Skoe E, Kraus N, Ashley R (2009) Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 29(18):5832–5840.
- Lehmann A, Schönwiesner M (2014) Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS ONE* 9(1):e85442.
- Lehmann A, Skoe E, Moreau P, Peretz I, Kraus N (2015) Impairments in musical abilities reflected in the auditory brainstem: evidence from congenital amusia. *Eur J Neurosci* 42(1):1644–1650.
- Lengagne T, Aubin T, Lauga J, Jouventin P (1999) How do king penguins (*Aptenodytes patagonicus*) apply the mathematical theory of information to communicate in windy conditions? *Proc Biol Sci* 266(1429):1623.
- Li X, Jeng F-C (2011) Noise tolerance in human frequency-following responses to voice pitch. *J Acoust Soc Am* 129(1):EL21–26.
- Liberman AM (1957) Some Results of Research on Speech Perception. *The Journal of the Acoustical Society of America* 29(1):117–123.

- Lieberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54(5):358–368.
- Lieberman AM, Mattingly IG (1985) The Motor Theory of Speech Perception Revised. *Cognition* 21(1):1–36.
- Lieberman AM (1996) *Speech: a special code* (MIT Press, Cambridge, Mass.).
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. *Cereb Cortex* 15(10):1621–1631.
- Lindstrom L, Klockhoff I, Svedberg A, Bergstrom K (1987) Abnormal auditory brain-stem responses in hallucinating schizophrenic patients. *Br J Psychiatry* 151:9–14.
- Liu F, Maggu AR, Lau JCY, Wong PCM (2014) Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Front Hum Neurosci* 8:1029.
- Malayeri S, Lotfi Y, Moossavi SA, Rostami R, Faghihzadeh S (2014) Brainstem response to speech and non-speech stimuli in children with learning problems. *Hear Res* 313:75–82.
- Manger PR, Pettigrew JD, Keast JR, Bauer A (1995) Nerve terminals of mucous gland electroreceptors in the platypus (*Ornithorhynchus anatinus*). *Proc Biol Sci* 260(1357):13–19.
- Marmel F, Parbery-Clark A, Skoe E, Nicol T, Kraus N (2011) Harmonic relationships influence auditory brainstem encoding of chords. *Neuroreport* 22(10):504–508.
- Marsh JT, Worden FG, Smith JC (1970) Auditory frequency-following response: neural or artifact? *Science* 169(3951):1222–1223.
- Mellado Lagarde MM, Drexel M, Lukashkina VA, Lukashkin AN, Russell IJ (2008) Outer hair cell somatic, not hair bundle, motility is the basis of the cochlear amplifier. *Nat Neurosci* 11(7):746–748.
- Møller AR, Jannetta PJ (1981) Compound action potentials recorded intracranially from the auditory nerve in man. *Exp Neurol* 74(3):862–874.
- Møller AR, Jannetta P, Bennett M, Møller MB (1981a) Intracranially recorded responses from the human auditory nerve: new insights into the origin of brain stem evoked potentials (BSEPs). *Electroencephalogr Clin Neurophysiol* 52(1):18–27.
- Møller AR, Jannetta PJ, Møller MB (1981b) Neural generators of brainstem evoked potentials. Results from human intracranial recordings. *Ann Otol Rhinol Laryngol* 90(6 Pt 1):591–596.
- Møller AR, Jannetta PJ (1982) Evoked potentials from the inferior colliculus in man. *Electroencephalogr Clin Neurophysiol* 53(6):612–620.
- Møller AR, Burgess J (1986) Neural generators of the brain-stem auditory evoked potentials (BAEPs) in the rhesus monkey. *Electroencephalogr Clin Neurophysiol* 65(5):361–372.
- Møller AR (2000) *Hearing: Its Physiology and Pathophysiology* (Academic Press).
- Moon IJ, Hong SH (2014) What is temporal fine structure and why is it important? *Korean J Audiol* 18(1):1–7.
- Moon IJ, Won JH, Park M-H, Ives DT, Nie K, Heinz MG, Lorenzi C, Rubinstein JT (2014) Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. *J Neurosci* 34(36):12145–12154.
- Moreno S, Bidelman GM (2013) Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear Res*. doi:10.1016/j.heares.2013.09.012.

- Muggenthaler E von (2000) Infrasonic and low-frequency vocalizations from Siberian and Bengal tigers. *The Journal of the Acoustical Society of America* 108(5):2541–2541.
- Musacchia G, Sams M, Nicol T, Kraus N (2006) Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res* 168(1-2):1–10.
- Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104(40):15894–15898.
- Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 241(1-2):34–42.
- Nagasawa M, et al. (2015) Social evolution. Oxytocin-gaze positive loop and the coevolution of human-dog bonds. *Science* 348(6232):333–336.
- Nan Y, Skoe E, Nicol T, Kraus N (2015) Auditory brainstem’s sensitivity to human voices. *Int J Psychophysiol* 95(3):333–337.
- Nourski KV, Brugge JF (2011) Representation of temporal sound features in the human auditory cortex. *Rev Neurosci* 22(2):187–203.
- Nuttall HE, Moore DR, Barry JG, Krumbholz K, de Boer J (2015) The influence of cochlear spectral processing on the timing and amplitude of the speech-evoked auditory brain stem response. *J Neurophysiol* 113(10):3683–3691.
- O’Beirne GA, Patuzzi RB (1999) Basic properties of the sound-evoked post-auricular muscle response (PAMR). *Hear Res* 138(1-2):115–132.
- Ouda L, Profant O, Syka J (2015) Age-related changes in the central auditory system. *Cell Tissue Res* 361(1):337–358.
- Parbery-Clark A, Skoe E, Kraus N (2009a) Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 29(45):14100–14107.
- Parbery-Clark A, Skoe E, Lam C, Kraus N (2009b) Musician enhancement for speech-in-noise. *Ear Hear* 30(6):653–661.
- Parbery-Clark A, Marmel F, Bair J, Kraus N (2011a) What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci* 33(3):549–557.
- Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N (2011b) Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6(5):e18082.
- Parbery-Clark A, Strait DL, Kraus N (2011c) Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* 49(12):3338–3345.
- Parbery-Clark A, Tierney A, Strait DL, Kraus N (2012a) Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219:111–119.
- Parbery-Clark A, Anderson S, Hittner E, Kraus N (2012b) Musical experience offsets age-related delays in neural timing. *Neurobiol Aging* 33(7):1483.e1–4.
- Parbery-Clark A, Anderson S, Hittner E, Kraus N (2012c) Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Front Aging Neurosci* 4:30.

- Parbery-Clark A, Strait DL, Hittner E, Kraus N (2013a) Musical training enhances neural processing of binaural sounds. *J Neurosci* 33(42):16741–16747.
- Parbery-Clark A, Anderson S, Kraus N (2013b) Musicians change their tune: how hearing loss alters the neural code. *Hear Res* 302:121–131.
- Parkkonen L, Fujiki N, Mäkelä JP (2009) Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Hum Brain Mapp* 30(6):1772–1782.
- Patterson F, Linden E (1981) *The education of Koko* (Holt Rinehart & Winston).
- Pettigrew JD, Manger PR, Fine SL (1998) The sensory world of the platypus. *Philos Trans R Soc Lond, B, Biol Sci* 353(1372):1199–1210.
- Plyler PN, Ananthanarayan AK (2001) Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. *J Am Acad Audiol* 12(10):523–533.
- Pratt H, Bleich N, Martin WH (1985) Three-channel Lissajous' trajectory of human auditory brain-stem evoked potentials. I. Normative measures. *Electroencephalogr Clin Neurophysiol* 61(6):530–538.
- Pratt H, Martin WH, Bleich N, Kaminer M, Har'El Z (1987) Application of the three-channel lissajous trajectory of auditory brainstem-evoked potentials to the question of generators. *Audiology* 26(3):188–196.
- Rana B, Barman A (2011) Correlation between speech-evoked auditory brainstem responses and transient evoked otoacoustic emissions. *J Laryngol Otol* 125(9):911–916.
- Richard C, Jeanvoine A, Veuillet E, Moulin A, Thai-Van H (2010) [Exploration of the auditory system in humans: From click to speech auditory brainstem responses]. *Neurophysiol Clin* 40(5-6):267–279.
- Ritsma RJ (1962) Existence Region of the Tonal Residue. I. *The Journal of the Acoustical Society of America* 34(9A):1224–1229.
- Rocha-Muniz CN, Befi-Lopes DM, Schochat E (2012) Investigation of auditory processing disorder and language impairment using the speech-evoked auditory brainstem response. *Hear Res* 294(1-2):143–152.
- Ruggles D, Bharadwaj H, Shinn-Cunningham BG (2012) Why middle-aged listeners have trouble hearing in everyday settings. *Curr Biol* 22(15):1417–1422.
- Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. *Clin Neurophysiol* 115(9):2021–2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N (2005) Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 156(1):95–103.
- Russo NM, et al. (2008) Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol* 119(8):1720–1731.
- Russo N, Nicol T, Trommer B, Zecker S, Kraus N (2009) Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Dev Sci* 12(4):557–567.
- Russo NM, Hornickel J, Nicol T, Zecker S, Kraus N (2010) Biological changes in auditory function following training in children with autism spectrum disorders. *Behav Brain Funct* 6:60.
- Sadeghian A, Dajani HR, Chan ADC (2011) Classification of English vowels using speech evoked potentials. *Conf Proc IEEE Eng Med Biol Soc* 2011:5000–5003.

- Scherg M, von Cramon D (1985) A new interpretation of the generators of BAEP waves I-V: results of a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol* 62(4):290–299.
- Shankweiler D, Fowler CA (2015) Seeking a reading machine for the blind and discovering the speech code. *Hist Psychol* 18(1):78–99.
- Shaw EAG (1974) The External Ear. Auditory System, *Handbook of Sensory Physiology.*, eds Keidel WD, Neff WD (Springer Berlin Heidelberg), pp 455–490.
- Shergill SS, Brammer MJ, Williams SC, Murray RM, McGuire PK (2000) Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Arch Gen Psychiatry* 57(11):1033–1038.
- Shinn-Cunningham B, Ruggles DR, Bharadwaj H (2013) How early aging and environment interact in everyday listening: from brainstem to behavior through modeling. *Adv Exp Med Biol* 787:501–510.
- Sinha SK, Basavaraj V (2010) Speech evoked auditory brainstem responses: a new tool to study brainstem encoding of speech sounds. *Indian J Otolaryngol Head Neck Surg* 62(4):395–399.
- Skoe E, Kraus N (2010a) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324.
- Skoe E, Kraus N (2010b) Hearing it again and again: on-line subcortical plasticity in humans. *PLoS ONE* 5(10):e13645.
- Skoe E, Nicol T, Kraus N (2011) Cross-phaseogram: objective neural index of speech sound differentiation. *J Neurosci Methods* 196(2):308–317.
- Skoe E, Kraus N (2012) A little goes a long way: how the adult brain is shaped by musical training in childhood. *J Neurosci* 32(34):11507–11510.
- Skoe E, Kraus N (2013) Musical training heightens auditory brainstem function during sensitive periods in development. *Front Psychol* 4:622.
- Skoe E, Krizman J, Spitzer E, Kraus N (2013) The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience* 243:104–114.
- Slabu L, Grimm S, Escera C (2012) Novelty detection in the human auditory brainstem. *J Neurosci* 32(4):1447–1452.
- Smith JC, Marsh JT, Brown WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39(5):465–472.
- Sohmer H, Pratt H (1977) Identification and separation of acoustic frequency following responses (FFRS) in man. *Electroencephalogr Clin Neurophysiol* 42(4):493–500.
- Sohmer H, Pratt H, Kinarti R (1977) Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42(5):656–664.
- Song JH, Banai K, Russo NM, Kraus N (2006) On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurotol* 11(4):233–241.
- Song JH, Banai K, Kraus N (2008a) Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiol Neurotol* 13(5):335–344.
- Song JH, Skoe E, Wong PCM, Kraus N (2008b) Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20(10):1892–1902.

- Song JH, Nicol T, Kraus N (2011a) Test-retest reliability of the speech-evoked auditory brainstem response. *Clin Neurophysiol* 122(2):346–355.
- Song JH, Skoe E, Banai K, Kraus N (2011b) Perception of speech in noise: neural correlates. *J Cogn Neurosci* 23(9):2268–2279.
- Song JH, Skoe E, Banai K, Kraus N (2012) Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex* 22(5):1180–1190.
- Spicer SS, Schulte BA (2005) Pathologic changes of presbycusis begin in secondary processes and spread to primary processes of strial marginal cells. *Hear Res* 205(1-2):225–240.
- Stapells DR, Oates P (1997) Estimation of the pure-tone audiogram by the auditory brainstem response: a review. *Audiol Neurotol* 2(5):257–280.
- Starr A, Achor J (1975) Auditory brain stem responses in neurological disease. *Arch Neurol* 32(11):761–768.
- Steinschneider M, Arezzo J, Vaughan HG Jr (1980) Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Res* 198(1):75–84.
- Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995a) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res* 674(1):147–152.
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG Jr (1995b) Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang* 48(3):326–340.
- Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA 3rd (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82(5):2346–2357.
- Steinschneider M, Nourski KV, Fishman YI (2013) Representation of speech in human auditory cortex: is it special? *Hear Res* 305:57–73.
- Stockard JJ, Rossiter VS (1977) Clinical and pathologic correlates of brain stem auditory response abnormalities. *Neurology* 27(4):316–325.
- Strait DL, Kraus N, Skoe E, Ashley R (2009a) Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 29(3):661–668.
- Strait DL, Kraus N, Skoe E, Ashley R (2009b) Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Ann N Y Acad Sci* 1169:209–213.
- Strait DL, Kraus N, Parbery-Clark A, Ashley R (2010) Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res* 261(1-2):22–29.
- Strait D, Kraus N (2011) Playing Music for a Smarter Ear: Cognitive, Perceptual and Neurobiological Evidence. *Music Percept* 29(2):133–146.
- Strait DL, Hornickel J, Kraus N (2011) Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav Brain Funct* 7:44.
- Strait DL, Chan K, Ashley R, Kraus N (2012a) Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex* 48(3):360–362.

- Strait DL, Parbery-Clark A, Hittner E, Kraus N (2012b) Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang* 123(3):191–201.
- Strait DL, Parbery-Clark A, O’Connell S, Kraus N (2013a) Biological impact of preschool music classes on processing speech in noise. *Dev Cogn Neurosci* 6C:51–60.
- Strait DL, O’Connell S, Parbery-Clark A, Kraus N (2013b) Musicians’ enhanced neural differentiation of speech sounds arises early in life: Developmental evidence from ages 3 to 30. *Cereb Cortex*. doi:10.1093/cercor/bht103.
- Strait DL, Kraus N (2014) Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear Res* 308:109–121.
- Swaminathan J, Krishnan A, Gandour JT (2008) Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19(11):1163–1167.
- Tarasenko MA, Swerdlow NR, Makeig S, Braff DL, Light GA (2014) The auditory brain-stem response to complex sounds: a potential biomarker for guiding treatment of psychosis. *Front Psychiatry* 5:142.
- Theunissen FE, Elie JE (2014) Neural processing of natural sounds. *Nat Rev Neurosci* 15(6):355–366.
- Tierney A, Parbery-Clark A, Skoe E, Kraus N (2011) Frequency-dependent effects of background noise on subcortical response timing. *Hear Res* 282(1-2):145–150.
- Tierney A, Kraus N (2013a) Neural responses to sounds presented on and off the beat of ecologically valid music. *Front Syst Neurosci* 7:14.
- Tierney A, Kraus N (2013b) The ability to move to a beat is linked to the consistency of neural responses to sound. *J Neurosci* 33(38):14981–14988.
- Troubetzkoy NS (1939) [Principles of Phonology]. Jednota ceskoslovenskych matematiku a fysiku (Prague).
- Tzounopoulos T, Kraus N (2009) Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62(4):463–469.
- Vedral V (2010) *Decoding Reality: The Universe as Quantum Information* (OUP Oxford).
- Wada SI, Starr A (1983a) Generation of auditory brain stem responses (ABRs). I. Effects of injection of a local anesthetic (procaine HCl) into the trapezoid body of guinea pigs and cat. *Electroencephalogr Clin Neurophysiol* 56(4):326–339.
- Wada SI, Starr A (1983b) Generation of auditory brain stem responses (ABRs). II. Effects of surgical section of the trapezoid body on the ABR in guinea pigs and cat. *Electroencephalogr Clin Neurophysiol* 56(4):340–351.
- Wada SI, Starr A (1983c) Generation of auditory brain stem responses (ABRs). III. Effects of lesions of the superior olive, lateral lemniscus and inferior colliculus on the ABR in guinea pig. *Electroencephalogr Clin Neurophysiol* 56(4):352–366.
- Wang JQ, Nicol T, Skoe E, Sams M, Kraus N (2010) Emotion and the auditory brainstem response to speech. *Neurosci Lett* 469(3):319–323.
- Warren RM (1968) Verbal transformation effect and auditory perceptual mechanisms. *Psychol Bull* 70(4):261–270.
- Weiss MW, Bidelman GM (2015) Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. *J Neurosci* 35(4):1687–1691.
- Wever EG, Bray CW (1930) Auditory Nerve Impulses. *Science* 71(1834):215.

- White-Schwoch T, Carr KW, Anderson S, Strait DL, Kraus N (2013) Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *J Neurosci* 33(45):17667–17674.
- White-Schwoch T, et al. (2015) Auditory-neurophysiological responses to speech during early childhood: Effects of background noise. *Hear Res* 328:34–47.
- Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 67(3):299–317.
- Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 128(Pt 2):417–423.
- Wiener FM, Ross DA (1946) The Pressure Distribution in the Auditory Canal in a Progressive Sound Field. *The Journal of the Acoustical Society of America* 18(2):401–408.
- Wiener N (1961) *Cybernetics or Control and Communication in the Animal and the Machine* (MIT Press).
- Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10(4):420–422.
- Worden FG, Marsh JT (1968) Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol* 25(1):42–52.
- Xing XY, Alström P, Yang XJ, Lei FM (2013) Recent northward range expansion promotes song evolution in a passerine bird, the Light-vented Bulbul. *J Evol Biol* 26(4):867–877.
- Xu Y, Krishnan A, Gandour JT (2006) Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 17(15):1601–1605.
- Yamada O, Marsh RR, Potsic WP (1980) Generators of the frequency-following response in the guinea pig. *Otolaryngol Head Neck Surg* 88(5):613–618.
- Yamada O, Marsh RR, Handler SD (1982) Contributing generator of frequency-following response in man. *Scand Audiol* 11(1):53–56.
- Yost WA (1996) Pitch of iterated rippled noise. *J Acoust Soc Am* 100(1):511–518.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256(5058):846–849.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11(10):946–953.